GENETIC PSYCHOLOGY MONOGRAPHS

Child Behavior, Animal Behavior, and Comparative Psychology

EDITED BY

CARL MURCHISON

VOLUME IX

1 9 3 1

Issued monthly by the
CLARK UNIVERSITY PRESS
Worcester, Massachusetts
U. S. A.

TABLE OF CONTENTS

$7.00 per volume
Single numbers $2.00
This issue $3.00

MONTHLY
Two volumes per year

Jan.-Feb., 1931
Volume IX, Nos. 1-2

GENETIC PSYCHOLOGY MONOGRAPHS

Child Behavior, Animal Behavior, and Comparative Psychology

THE STATUS OF THE FIRST-BORN WITH SPECIAL REFERENCE TO INTELLIGENCE*

From the Institute of Child Welfare, University of California

By

HSIAO HUNG HSIAO

RESEARCH FELLOW OF THE CHINA FOUNDATION FOR THE PROMOTION OF EDUCATION AND CULTURE

*Accepted for publication by Harold E. Jones of the Editorial Board and received in the Editorial Office October 6, 1930.

Worcester, Massachusetts

Entered as second-class matter December 1, 1925, at the post-office at Worcester, Mass., under Act of March 3, 1879

ACKNOWLEDGMENTS

The writer is indebted to Professor Harold E. Jones for the guidance of this research and for administering the greater part of the data collection. Acknowledgments are also due to Dr. Virgil Dickson, Assistant Superintendent of the Berkeley School Department, for his generous aid in connection with the assembly of test records.

HSIAO HUNG HSIAO

HUNAN UNIVERSITY
CHANGSHA, HUNAN, CHINA

CONTENTS

I

INTRODUCTION

The Origin of the Problem

A survey of the literature on the relationship between the order of birth and the mental and physical traits of the child leads to the realization that there is lack of general agreement concerning the comparative status of the first-born. While they are said to be liable to physical handicap of many sorts, they are nevertheless credited with exceptional longevity. They are found to predominate among geniuses and scientists, and also among idiots and imbeciles. In view of numerous apparent contradictions in the reported data, it seems desirable at this time to make a critical analysis of previous investigations and to re-examine the problem in certain of its most important phases.

The Significance of the Problem

The significance of the problem may be gleaned from the following facts: Historically, we have seen that the differential effect of birth order manifested itself in various aspects of life long before this question was subjected to a scientific analysis. For instance, the priority of birth often entitles the man to many special privileges, including those of family leadership, religious priesthood, and property inheritance. Royal heirships have also been chiefly based upon the same principle. It is within the province of anthropology to decide whether these practices are merely a matter of socio-legal convenience, or ultimately traceable to a

folk-belief in the superiority of the first-born [like that which prevails in northern Iceland (24)]. The present study is concerned with determining whether or not there is a bio-psychological justification for this differential function of birth order. In view of the disagreement that exists among the findings so far available, the latter question has not yet been completely solved.

There is another aspect of the problem to be mentioned, one which derives its significance from the Malthusian practice. It has been clearly brought out in the following remarks of Pearson (96), whose findings are in favor of the later-born: "It means that if you reduce the size of the family, you will tend to decrease the relative proportions of the mentally and physically sound in the country. You will not upset this conclusion, if, as I suspect, the extremely able man, the genius, is also among the early born. For you will not lose him if you have a larger family, although you will lose the sounder members if you curtail it." These statements serve to indicate the implications of this problem in eugenics.

From the above, it may be readily seen that, from a legal, a sociological, and a eugenical point of view, the problem is a highly significant one, while at the same time it is not lacking in a purely psychological and biological interest.

THE GENERAL CONSIDERATIONS NEEDED IN THE SOLUTION OF THE PROBLEM

Superficially, the problem may seem to be merely one of comparing different birth orders, but, in fact,

it is this very deceptive appearance of simplicity which has constituted the chief source of error in prior studies. It is clearly shown in the chapter on order of birth in Holmes' book, *The Trend of the Race* (70), that the problem is beset with numerous statistical and other pitfalls. Therefore, in attacking the problem, one will do well to be on the lookout for the various fallacies the results are likely to involve. It seems to the writer that these pitfalls or fallacies are preventable, or at least reducible to a minimum by paying due attention to certain considerations. This section will be devoted to the discussion of general considerations, and the more specific ones will be taken up in later chapters where some of the previous investigations may serve as illustrations.

1. In the first place, we have to consider whether the study is to deal with the socially or biologically first child. If the latter is to be our concern, then the previous prenatal casualties (abortions, miscarriages, and still-births) will have to be taken account of, otherwise the so-called first-born might belong to any birth position and therefore would fail to represent the true status of primogeniture. But this procedure of determining the genuineness of the first birth is not always possible, and, in such cases, it will be desirable to resort to some means of determining the probable incidence of prenatal casualties in the sample studied, in order to know how the results are thereby affected.

2. The next thing to be considered is the incompleteness of families. If the study is concerned with the relative frequencies (such as percentages, permilles, etc.) of various birth orders in the material, it will be

necessary to make sure that the data include only completed families. The importance of this point may be made evident in the following manner: Being computed on the basis of the total data collected, the relative frequencies of different birth orders will naturally vary with the number of the later-born included in the data. For instance, 200 first-born, in a material containing only 600 later-born, will be 25 per cent of the total material, which consists of 800 cases, but the same number of first-born, in a material including 800 later-born, will be only 20 per cent of the total material, for the latter now contains 1000 individuals. Therefore, the larger the number of incomplete families included in the material, the smaller the number of the later-born and the higher the relative frequencies of the earlier-born. Now, if the relative incidence of a certain trait among different birth orders constitutes the object of study, the inclusion of incomplete families is likely to leave out a number of younger children who are yet to be born and who will possess the trait in question, and consequently the relative frequencies of the earlier-born in the material will be exaggerated. It should also be noted that, if the problem is to compare the eldest and the youngest in terms of average degrees of a trait, the presence of incomplete families in the data will also seriously affect the accuracy of results, for the youngest children at the time of investigation will not necessarily remain to be such afterwards.

3. It is an easily recognizable fact that the relative frequencies of various birth orders found to be associated with a certain trait will vary with their rela-

tive frequencies in the general population if the trait in question is evenly distributed among the different birth orders. This is because the larger the number of certain birth orders in the general population the more frequently they will be observed. Such being the case, the mere preponderance of certain birth orders in the data collected does not necessarily signify that these birth orders are actually weighted by the trait in question, for their relative frequencies in the material are always affected by their relative frequencies in the general population. However, the latter are in turn influenced by the following factors:

a. The rate of mortality. The rate of mortality may be selective of certain birth orders, and it may be differential for different birth orders at different ages. The relative frequencies of different birth orders in the general population may vary accordingly. For instance, the proportion of the first-born among the children will be small if they suffer from a higher death rate than the rest of the birth ranks during the period of infancy, but their proportion among the adults will be large if, from the age of five to the age of twenty-five, their rate of mortality is to a sufficient degree exceeded by that of the other birth orders. These conditions will materially influence the data collected, and therefore the relative frequencies of the first-born in the material cannot be expected to be constant at different ages unless their rate of mortality remains in a constant ratio to that of the later-born at these ages.

b. The rate of birth. A secular trend in the annual rate of birth may also cause fluctuations in the proportion of the first-born in the general population,

and these in turn influence the collection of data. This situation has been well described by Cobb (29) as follows: "It has hitherto been assumed that if a person of given age is selected at random from amongst fraternities of a given size, then all positions in that fraternity are equally likely. But this is not the case. If the number of births has been increasing he is more likely to be one of the oldest members of his fraternity, and if the number of births has been decreasing he is more likely to be a younger member. For while the number of births is increasing there are more children born every year who belong to the first half of their fraternities than who belong to the second half."

c. The social composition of the community. The fertility and mortality of the population are partly conditioned by the social classes that compose it. Therefore, the relative frequencies of different birth orders will differ in the data drawn from different social strata, and will have to be evaluated with due regard for the peculiarities of the latter. For instance, if one is to obtain data to serve as the basis of comparison for the birth order distribution in the material, it is imperative that these be drawn from the very community from which the material is derived or from a community of similar social composition, otherwise the results will be misleading.

4. The size of family may influence the results in a variety of ways. Certain traits may be frequently associated with small families and others with large families. If a trait of the former sort is to be studied, the material is likely to include a large proportion of small families, and consequently the first-born will ap-

pear to be weighted. The reverse condition will prevail if the selection is in favor of large families. Therefore, it is necessary to distinguish between the weighting of families and the weighting of birth orders.

However, it is not always a safe procedure to restrict the data to families of a constant size. This point has been brought to light by Holmes (70, pp. 306-307), who may be quoted as follows: "Suppose that of the families starting in 1825, one in ten contained a twelfth child, which lived for 60 years. Suppose also that of the families starting in 1875 only one in one hundred had a twelfth child that lived for 35 or more years. Now suppose that in 1910 we select a group of individuals from the families of twelve in the population. Our group would contain many more of the twelfth born in the older families than from the later ones. When we simplify the problem of handicapping the first-born by considering the ordinal position of the marked members within families of a given size, we do not avoid all statistical pitfalls."

Moreover, certain traits may differ in magnitude in families of different sizes, and therefore comparisons of different birth orders in terms of average degrees of such traits will be in favor of the earlier- or later-born, according as the magnitude of the traits varies inversely or positively with the size of the family. Take intelligence as an example. Intelligence has frequently been reported to vary inversely with the size of family (16; 26; 32; 79; 83; 119; 120, pp. 121-125). In other words, the larger the family, the lower the intelligence of its members is likely to be. Now, if averages are taken of different birth orders without regard to the

size of the family, the results are apt to be in favor of the earlier-born, for the children of higher birth ranks are necessarily drawn from larger families. On the other hand, aside from the artificial restriction of family size, small families may be indicative of the lower vitality of parents. These small families, if mixed with those that have resulted from intelligent control, may partly or wholly offset the advantage derived from the latter.

5. The age of the mother at the birth of the child has been shown to have an effect upon certain of his traits (9, 11, 40, 41, 65, 86, 105, 117, 124, 125), particularly in the case of extremely young mothers, or mothers near the end of the child-bearing period. Therefore, comparisons of different birth orders in a trait in terms of its average degrees without consideration for this factor will leave open the possibility that the differences found are due to the differences in the mother's age. But failure to find differences among the birth orders is not attributable to the mother's age unless there is likelihood that this factor operates to cancel some other factor which is believed to act in an opposed direction.

6. The social and economic status of the family is undoubtedly an important factor in determining some of the traits of children. It is easily conceivable that if the material is heterogeneous in its social composition, it is likely to create spurious differences among birth orders if averages are taken without controlling the social and economic status of the family. Furthermore, there are certain factors which may have a greater differential effect upon the birth orders in the

case of certain social classes than in the case of others, such as the lack or inadequacy of provision for the intellectual stimulation of the first-born during early childhood, the possibility of bestowing greater care upon the younger child at the expense of the elder, etc. In view of this fact, birth order differences that are found in an inferior community but not in a superior one may be attributable at least partly to social and economic factors.

7. The possible bearings of the incidence of syphilis upon the results of this problem have been pointed out by Holmes (70, pp. 306-307). He states that the common characteristic of a syphilitic family is first the occurrence of some abortions, then the birth of sickly children, and finally the production of children who are comparatively healthy. If this proves to be the case, the first-born will tend to occupy an unfavorable position in any material that includes a considerable number of such families.

Summary

The problem of the handicapping of the first-born has constituted the object of many studies, but, in view of the disagreement of the findings, no definite conclusions can, as yet, be reached.

It has been shown that the problem possesses a legal, a sociological, and a eugenical significance apart from the purely psychological and biological interest to which it is entitled.

Special attention has been called to certain general considerations bearing upon the problem. Among them were included such questions as the socially ver-

sus the biologically first-child, and the influence of incompleteness in families. It has also been pointed out that the relative frequencies of different birth orders in the data collected will vary with their relative frequencies in the general population, and that the latter, in turn, will vary with the rate of mortality, the rate of birth, and the social composition of the community. The other considerations that have been taken up in this chapter are the size of the family, the age of the mother, the social and economic status of the family, and the syphilitic history of the parents. An attempt has been made to show how these factors may influence statistical results.

II

HISTORICAL RETROSPECT

It is worthy of note that, in the last sixty years, interest in this problem has steadily grown, and that there has appeared in various countries a considerable number of studies concerned with this problem. These will be reviewed for the double purpose of clarifying some of the issues involved and of indicating the places where a restudy of the problem is needed. For the sake of clearness, these studies will be divided into three categories: those that deal with physical traits, those that deal with behavior traits, and those that are not classifiable under either of these two. Under each category sub-divisions have been made. The studies that belong to each of these will be presented in chronological order except where other considerations are involved.

It may also be mentioned that, in this review, studies of only children will not be included. This is for the following reasons: (*a*) only children may belong to any birth order, and (*b*) the only child situation is one which involves factors sufficiently complicated to require an independent treatment.

A. Physical Traits

1. *Height and Weight.* As early as 1864, Duncan (37), investigating 2087 infants in the Edinburgh Royal Maternity Hospital, found that primiparous mothers had children averaging 7.17 pounds in weight and 19.2 inches in length, while multiparous mothers

had children averaging 7.27 pounds in weight and 19.2 inches in length. Weight increased with the age of the mother up to the period of 25 to 29 years and then decreased. This led him to believe that the age of the mother and not the order of pregnancy was the significant factor.

Hecker (63) measured 4449 cases in a maternity home in Berlin, and reported that the weight of the infant increased with the mother's age up to the age of 44 years, and also increased with the order of birth. He was inclined to think that birth order was the more important factor.

Castell (24), reviewing the reports of Duncan and Hecker and studying a considerable amount of additional material, concluded in favor of Hecker's view that birth order rather than mother's age should account for the phenomena found.

Fasbender (43) investigated 630 cases at a maternity home in Stuttgart. He found that the largest children were those born to mothers between 25 and 35 years old, and that children of primiparae were lighter and shorter than those of multiparae.

Issmer (75), in an original study of length and weight measurements of 12,303 infants in a maternity hospital in Dresden, considered that the most significant characteristics of a newborn child were length and weight. He found that the first-born children were lighter than those of subsequent birth.

Schmid (112) studied 2700 well children whose ages ranged from one day to thirty months. All the older children had been breast-fed for at least nine months. According to his findings, the children of old

pluriparae were at birth heavier than those of young primiparae, but this relation was reversed after the first year.

Fourmann (44) reported on over 1000 cases to the effect that birth order was positively correlated with birth weight. He also found that the duration of pregnancy was correlated with birth weight, and that the age of the mother had no considerable effect. The influence of duration of pregnancy was what might be expected, but it leads to the query as to the possibility of differences in the average duration of pregnancy among primiparae compared with multiparae. If we should find children of the former born younger in terms of conceptional age, this alone might explain their inferior height and weight.

Révész (106), in his study of the influence of the mother's age upon the stature of her child, observed that the youngest children were often the tallest, but showed, on the basis of both his personal experience and anthropological data, that mother's age was a determining factor.

H. J. Hansen was quoted by S. Hansen (61) as showing that the weight of newborn children increased with birth number up to the fourth or fifth child or over, and that the weight of the mother at birth was a factor of minor importance. The difference between the average weight of the first and fourth child, when the age of the mother was controlled, was not less than 300 grams, which corresponds to a difference of 6 kilos for adults with an average weight of 70 kilos.

Peller (98) studied the records of 5487 newborn

infants obtained from two maternity institutions in Vienna to determine the effect of social conditions on the physical development (especially the weight) of newborn infants. The data used were based upon 5026 full-term infants whose mothers were of various social classes and economic conditions. The average weight of the first-born children in one institution was 3255 grams for the boys and 3145 grams for the girls. The figures increased with each successive birth until they reached, for the fifth to the seventh birth, 3515 grams for the boys and 3361 grams for the girls. In the other institution the figures were 3376 grams and 3237 grams for the first-born boys and girls, respectively, and 3608 grams and 3452 grams for the boys and the girls of the fifth to the seventh birth. The author concludes that the average weight increases with the number of pregnancies, but the difference tends to become smaller.

Harris (62) reported results for six racial groups. Pregnancy order and birth order were distinguished, the former taking account of miscarriages. There was a small positive correlation between pregnancy order and weight of child, and also between birth order and weight of child. The coefficients were from +.14 to +.21. When length was correlated with birth order and with pregnancy order, it proved to be in both cases slightly positive for males, but slightly negative for females.

Pearson (94) made a comparison of 385 brother-brother pairs and 450 sister-sister pairs, and found that the younger siblings were taller than older ones. Further evidence along this line has been shown in his study of 2000 babies. Excluding twins, and confining

attention to legitimate and normal-time infants, the results shown in Table 1 were obtained. Here the first-born have less weight for both boys and girls than any children subsequently born. Precisely the same point is brought to light in Table 2. However, contrary

TABLE 1

WEIGHT OF NEWLY BORN BABIES IN ORDER OF BIRTH CATEGORIES

(Weight in pounds)

	Birth order 1	2	3-4	5-6	7-8	9-10	11 and over	Total cases	Mean weight
Boys	7.01	7.36	7.41	7.70	7.91	7.59	7.92	856	7.40
Girls	6.76	7.08	7.33	7.36	7.32	7.65	7.88	866	7.15

TABLE 2

LENGTHS OF NEWLY BORN BABIES IN ORDER OF BIRTH CATEGORIES

	Birth order 1	2	3-4	5-6	7-8	9-10	11 and over	Total cases	Mean weight
Boys	20.62	20.82	20.80	20.95	20.98	20.99	20.14	856	20.81
Girls	20.27	20.33	20.51	20.43	20.36	20.41	20.73	866	20.38

findings are not lacking. Boas (14) found that the first-born children exceeded the later-born in stature as well as in weight, in an age range from the sixth year to the adult state in females and from the sixth year to the fifteenth year in males. The differences in stature were 10 mm. in girls and 7 mm. in boys; and the differences in weight were 1.6 pounds in girls and 1.2 pounds in boys. The author was of the opinion that the cause of the phenomenon might be the greater weight of the mother at the time of birth of the first child and the greater care bestowed upon the first child during its early childhood.

Wells (131) has been quoted by both H. Ellis and W. C. Rivers, as shown in an unpublished thesis, that the eldest children tended to excel in both stature and weight. Youngest children, while inferior to eldest, were superior to intermediate children.

Ewart (41) has found that "the first and second do not differ very much in stature, and that the fall begins really from the third onwards." The subjects selected for the study were children in their sixth and thirteenth years.

Reiter and Ihlefeldt (105) studied 634 legitimate children born in Rostock, and found that the later-born were inferior in size and weight. According to the author, "Mehrgebürtigkeit" was responsible for the phenomenon; for, in those cases, the duration of nursing was found to have no significance.

At the first sight, there seems to be lack of agreement among these findings. But, on closer inspection, one will find that the significant findings in favor of the later-born are those dealing with newborn infants, while the weighty evidences in favor of the earlier-born are those concerned with older children. Therefore, the seeming contradiction does not necessarily exist.

In this connection, there are several other investigations which might be mentioned. Chase (27) studied 58 sets of brothers who were college students. According to his results, there is no birth order significance as far as height and weight are concerned. The rest are the studies chiefly concerned with the loss or recovery of weight. Kézmárszky (82) studied 73 normal breast-fed children, and found that the first-born chil-

dren made a slower recovery from the physiological weight loss occurring immediately after birth. Mourlot (90), on the basis of 125 infants born at term; Griffith, Crozer, and Gittings (60), on the basis of 126 full-term infants; and Adair (2), on the basis of 298 children, all concluded that the initial loss of weight was less in infants of multiparae than those of primiparae. Hillebrand (67) found that the first-born infants were approximately 1½ days behind others in weight and in appetite.

2. *Skeletal and Muscular Conditions.* Cadre (22) investigated the condition of the sagittal suture of 203 infants from birth to nine days of age. He found that the overlapping of the sagittal suture, due presumably to compression during delivery, was most frequent among the infants of primiparae.

According to the findings of Jansen (76), there is an increasing failure of proper skeletal development in the successive children in large families.

In the above-mentioned study by Chase (27) it was found that the eldest sons were the weakest on strength tests.

3. *Diseases and Physical Peculiarities.* Brehmer (18) studied 494 cases of consumption. His general impression was that, in certain instances, scrofula affected earlier offspring; and that, in large families, the later-born were likely to be consumptive, especially the sixth. However, his data were afterwards analyzed by Rivers (108), who reported a distinct excess of phthisis among first- and second-born.

In the same report Rivers (108) made an analysis of Riffel's three studies: "Schwindsucht und Krebs im

Licht der vergleichend statistisch-genealogischen Forschung," "Weitere pathogenetische Studien über Schwindsucht und Krebs," and "Mitteilungen über die Erblichkeit und Infektiosität der Schwindsucht." These studies, for the sake of brevity, may be designated as *S.K.*, *W.P.S.*, and *E.I.*, respectively. He was aware of the fallacy that, owing to a low infant mortality, the adolescent and adult population would contain an unduly large population of eldest-born, and, consumption being a disease mainly of these ages, consumptives would show an unduly large proportion of eldest-born likewise. To guard himself against this fallacy, he limited the material to those who had died at adolescent or adult stage, and the result was as follows: In the *S.K.* material the first-born deaths from phthisis constituted 24.1% of the deaths that had occurred at 18 years and upwards. This percentage was the highest except in the case of the twelfth-born, whose number was small. In *W.P.S.* the first-born proportion was also the highest of all. In *E.I.*, the effect of such an analysis was to bring the percentages of deaths from phthisis among the first-born above the general average.

This analysis was followed by a study of the records of sanatorium patients. They related to sputum-positive cases only. As in Brehmer's material, there was considerable excess of consumption among the first- and second-born. There was also excess among the fifth-born. When the cases of consumption among the patients' brothers and sisters were added, the results were practically the same. If, to exclude a possible fallacy before-mentioned, the distribution of consumption in order of birth was taken among subjects only of

18 years of age and upwards, there was still considerable eldest-born excess (actual 75, expected 60.5) and some second-born excess.

Pearson (95, 97) also found heavy incidence of tuberculosis on the first- and second-born among 2164 members of 381 families. The numbers of cases observed for the two birth ranks were 113 and 79, and those expected were 67.1 and 64.4, respectively. In order to give additional confirmation to this investigation, he took 7670 fertile families from the industrial class in New South Wales, which were completed at the time of the record. The total children were 46,325 and the average size of family was 6.05, which was only slightly in excess of that of the tuberculous stocks. According to Pearson, the general correspondence between the distribution of elder and younger siblings in New South Wales' industrial classes and that in the tuberculous families was sufficient to indicate that the great excess of elder members among those actually suffering from tuberculosis was not due to any peculiar birth order distribution of the tuberculous families.

The conclusion drawn by Pearson that the eldest-born is generally inferior elicited a reply from Von der Velden (126), who took the tuberculous families in Riffel's *W.P.S.* and *S.K.*, and made out that while they contained more deaths from consumption among the first three birth ranks than in the remainder, yet this was due to lower infantile mortality. A similar criticism was made by Weinberg in his paper "Die rassenhygienische Bedeutung der Fruchtbarkeit" (128). But, in a previous paragraph, we have seen how this point was tested by Rivers.

In "Die Kinder der Tuberkulösen" (129), on the basis of a study of the children of the tuberculous, Weinberg concludes: "At any rate it appears from my investigation that the early children, in consequence of their lower total mortality in childhood up to the 20th year, are more predominantly represented among the adults than among the total number of births. Now, as the material on which Pearson and Rivers have based their views consists of the anamneses in the sanatorium; therefore, instead of representing the total population, it refers chiefly to adults whose birth order distribution is compared by these authors with that of the total number of children from the same families and not with the distribution of adults in general. From what has been said, it is perfectly clear that, with a comparison of that kind, the earlier-born will, under all circumstances, come out poorly even if they do not actually in the same age contract tuberculosis or die of it more frequently than the later-born. Therefore, we may accept the results of the above-mentioned authors only provisionally; i.e., as results of an incorrect method, until evidence has been found to the contrary through a more exact formulation of the problem."

Another attack upon Pearson comes from Macauley (84). We may quote his own words: "In Dr. Pearson's tables the inclusion of children and babies who are not yet old enough to be exposed to the dangers in question, very clearly makes it inevitable that the groups of later-born, to which these children belong, will show a more favorable percentage than the groups of early-born who are of necessity grown-up. To draw any serious deductions from such sta-

tistics would on its face appear absurd. Dr. Pearson, however, concludes that there are as many people dead, or who have passed the supposed danger zone, as there are children in the families not yet old enough to be exposed . . . . . it appears to be in direct conflict with the statistics of the disease.

"To those who think the assumption is sound, we commend the following statistics from the report of the Registrar General of England and Wales, 1891-1900.

TABLE 3

ANNUAL MORTALITY FROM PHTHISIS PER MILLION LIVING AT VARIOUS AGES

Age	Mortality per million	Age	Mortality per million
0 to 4	413	35 to 44	2592
5 to 9	206	45 to 54	2362
10 to 14	368	55 to 64	1881
15 to 19	1144	65 to 74	1154
20 to 24	1730	75 and over	437
25 to 34	2135	All ages	1391

"In the face of such figures it is difficult to see how the claim can be supported that the older members are 'past the danger zone,' and therefore, not likely to become inmates of a sanatorium. Not until age 75 does the death rate from phthisis descend to the neighborhood of what it is below age 15. The professor claims that the mean age of onset of the disease is 29.1 for males and 25.3 for females. That, however, is merely another way of saying that at those average ages the patients began to be eligible for admission to the sanatorium. Patients may be admitted not merely at the onset of the disease, but probably at any time prior to death. Not merely, therefore, is the assumption which

we are asked to accept entirely unproved, but in view of such statistics as the above appears to be entirely disproved" (84, p. 169).

S. Hansen (61) has also made a study of the disease, the material for which is obtained from Oresund Hospital and Boserup Sanatorium. According to him, the Boserup material is better than that of the Oresund Hospital, because the former includes only absolutely sure cases of pulmonary tuberculosis, while the latter contains a number of doubtful cases although this number is not large enough to influence the results seriously.

The Oresund material includes 1887 men and 1635 women. The result of the investigation shows that not only the first-born but also the second and third children are more frequently attacked by tuberculosis than they would be were their places in the order of sibs without significance. This result is in full agreement with Pearson's. If the study is limited to families of equal size—say five sibs—the predominance of the earlier born (the first and second) comes out even more clearly.

In the Boserup material, families of five suffer from irregularities, and seem to indicate that the third child is more frequently tuberculous than the rest; but those of four show a clear excess of the first child. In families of six, there are 141 tuberculous patients in the older half against 129 in the younger half; and, in families of eight, 114 against 83. However, these two groups are too small to have any decisive significance.

To further confirm the findings, a comparison has been made of the relative frequency of the first-born among tuberculous patients in the Oresund material

and in the marriages of 25 years' duration or more. This limitation to marriages of 25 years' duration or more, as the author says, is based upon the fact that the average age of tuberculous patients is about 30 years. The results are shown in Table 4.

TABLE 4
BIRTH ORDER OF TUBERCULOUS PATIENTS COMPARED WITH THAT OF CENSUS*

No. of birth	Census	T.B. families	T.B. patients	No. of birth	Census	T.B. families	T.B. patients
1	173	171	281	9	35	31	13
2	159	162	202	10	23	22	14
3	143	148	161	11	15	14	7
4	124	127	121	12	9	10	7
5	103	105	77	13	6	6	2
6	85	84	56	14	3	4	1
7	66	63	32	over 14	6	7	3
8	50	47	23	Totals	1000	1000	1000

*The Census refers to the number per mille according to the census of February 1st, 1901, found in Cordt Traps' Bearbeitung der Kopenhagener Ehestatistik (Cordt Trap: Kinderzahl und Kindersterblichkeit in der Kopenhagener Ehen. Kopenhagen, 1905).

The excess of the relative frequency of the first-born among the tuberculous patients over that both in the census and among the children of tuberculous families is regarded by Hansen as indicative of the heavy incidence of tuberculosis on the first-born.

The fact that tuberculosis appears so frequently among the first-born is usually explained on the assumption that they are older and therefore more exposed to the disease than their later-born sibs. In order to invalidate this assumption, the author has presented the age distribution of tuberculous patients of Boserup in a table, which is reproduced here as Table 5.

TABLE 5

Distribution of Ages According to Birth Order

Birth No.	Male patients No.	Male patients Age in years	Female patients No.	Female patients Age in years
1	297	29.1	209	28.3
2	208	29.5	200	27.4
3	194	30.1	167	27.1
4	134	29.3	121	27.3
5	99	30.7	88	27.4
6	82	29.9	55	26.5
7	50	30.7	48	29.5
8	22	30.4	37	26.8
9	24	30.9	15	26.3
10	13	29.5	7	24.6
Over 10	19	31.2	24	25.8
Total	1142	29.6	971	27.4

On the basis of this table, the author concludes: "The earlier-born men are even somewhat younger than their later-born sibs, whereas no difference exists with women; and therefore we can set aside the possibility that the age-difference which is present within individual families could be reflected in the total material" (61, p. 714).

However convincing it may appear to Hansen, the last argument seems far from satisfying to the reviewer. The latter is of the opinion that, in order to disprove the above-mentioned assumption, it is necessary to determine the presence or absence of a *significant* difference between the average age of the earlier-born and that of the later-born among all the children of tuberculous families. Unless such a difference is found to be actually absent, we are not justified in denying age as a selective factor in the excess of the earlier-born among the tuberculous patients; because, according to the assumption, the later-born, if found to be equal in age to

the first-born, should be equally exposed to the disease. As to the absence of age difference between the earlier-born and the later-born among the patients themselves, it signifies no more than that the disease tends to manifest itself more strongly around a certain age. This fact could just as well be used by Hansen's opponents to support their assumption by stating that, since the disease tends to manifest itself in adult age, it is therefore more likely that the earlier-born would be more exposed to the disease, inasmuch as they are older than their sibs. Hansen has undoubtedly overlooked this point.

Among the critics of Pearson we might also mention Greenwood and Yule (59). In criticizing Pearson's method of comparing the proportion of the first-born among marked individuals with the proportion of the first-born among their sibships, they have endeavored to show that "the proportion of rth born among the sibships of marked individuals is not the same as in the population at large." The mathematical basis of their argument is as follows (59, pp. 187-188):

"In any population in which the number of sibships of x members is fx, the number of rth born ($r^{S}x$) contributed by those sibships is fx; hence the proportion of rth born in the whole population is:

$$ {}_{a}B_r = \Sigma_r^k (fx) \Big/ \Sigma_1^k (xfx) $$

where k is the greatest size of sibship occurring.

"The proportion of rth born in the marked sibships, on the other hand, is, if we count each sibship only once:

$$ {}_{s}B_r = \Sigma_r^k (Fx) \Big/ \Sigma_1^k (XFx) :: $$

$$\Sigma_r^k \left[fx \left(1-q^x \right) \right] \Big/ \Sigma_1^k \left[Xfx \left(1-q^x \right) \right]$$

or if we count each sibship once for each marked member:

$${}_sB'_r = \Sigma_r^k \left(F'x \right) \Big/ \Sigma_1^k \left(XF'x \right) =$$

$$\Sigma_r^k \left(Xfx \right) \Big/ \Sigma_1^k \left(X^2fx \right)$$

"In neither case does the proportion of rth born amongst the marked sibships approximate to the proportion of rth born in the population at large; but is too small for the earlier-born, in consequence of the manner in which the larger sibships have been weighted.

"Hence if we compare the proportion of rth born among marked individuals with the proportion of rth born among their sibships, we are making a fallacious comparison, which must inevitably lead to an apparent incidence of the marking on the earlier-born."

Pearson has also been criticized by Dublin and Langman (35), but from a slightly different angle. They contend that Pearson's "method is based unequivocally on the assumption that the distribution according to order of birth of the pathological community from which his marked or affected subjects are obtained is identical with the distribution of the sibships of these subjects." They have tried to show that "when there is no weighting according to order of birth among the individuals affected, the distribution of the affected or that of the pathological community represented by them is not in any case comparable with that of their sibships."

However, the statistical results that Dublin and Langman have arrived at still show a preponderance of the first-born among the tuberculous, although the authors do not regard it as significant.

	Birth order												
	1	2	3	4	5	6	7	8	9	10	11	12	13
Actual no. cases	90	81	61	59	40	25	15	11	7	7	5	8	3
Expected no. cases	62	59	56	40	32	31	29	29	15	14	5	4	2

Here there was heavy incidence of epilepsies in the first three children. However, within the individual groups of equal-size families the difference failed to come out distinctly. In this connection attention may be called to Pearson's (97) conclusion that, while there is a weighting of the elder-born in epilepsy, this is due to a selection of families rather than to a selection of the elder-born in each individual family.

In Czellitzer's (30) material dealing with the heredity of high degree of short-sightedness (more than -6D), Auerbach (6) found that of 100 children from short-sighted families 34.6% were attacked by the same trouble, and of these 48% were first-born. However, the author has asserted that the concept of the first-born is a purely juristic, and not a "vererbungstechnischer," one. According to him, an explanation of Czellitzer's findings may be found in the fact that this high degree of short-sightedness tends to progress up to puberty. The first-born, being relatively the oldest of the children investigated, will naturally, in a larger proportion, exceed the arbitrarily established limit of -6D. With younger children either the progressing tendency of myopia is not yet established or the degree reached is not yet sufficient to make them classifiable among the high-degree short-sighted. In regard to this phenomenon, Weinberg (130) suspects the existence of some kind of clinical selection in it.

As to the question of the relationship between "in-

sanity" and birth order. Heron's (64) first study of the statistics of insanity is the only one which has come to the writer's attention. Heron found that the first-born children from insane stock were more likely to become insane than other members of the sibship. His tables show that this tendency may be observed in the first four siblings, but is especially marked in the case of the eldest, where the frequency was found to be 108 as against a chance expectancy of 77.3. The data are based upon 468 insane individuals coming from 315 families, with a total sibship of 1908. The same data were studied again by Pearson (97), by comparing the distribution of the actual insane with that of their offspring. The permilles of the four lowest birth numbers of the former were 231, 171, 167, and 152; while those of the latter were 193, 182, 153, and 133. Therefore, except in the case of the second-born, heavy incidence of insane was shown on the other three birth numbers. Here Pearson (97) has pointed out that 193 is "a maximum exaggerated by the incompleteness of the families, by the check which confinement in an asylum places on fertility, and by the fact that smaller families are more customary now than in the generation of the parents of the insane."

There are two other studies of Pearson (97) which may come under this category. One of these is a study of congenital cataract from pedigrees in his possession. The material suggests that this disease, which is markedly hereditary in character, has a bias against the two elder-born. The method employed is to compare the number of the affected members belonging to different birth ranks with the sibships of parents of the cataract

ous as well as with the siblings of the cataractous. The other study (97) is of 952 albinos. It is pointed out that there is marked excess of albinos among the first-born, that the second-born have no excess, and that there seems to be a defect among the last-born.

In a study of a neurotic inventory, L.L. and T.G. Thurstone (122, pp. 25-26) have briefly touched upon the question of the relation of birth order to personality development. The subjects were asked the following questions:

"Did you grow up as an only child?"
"Did you grow up as the youngest child?"
"Did you grow up as the oldest child?"
"Did you grow up as the next to the oldest child?"

The proportions of "yes" and "no" answers to these questions were about the same in the best-adjusted and neurotic groups, and, on the basis of this fact, the authors have concluded that the question of birth order is not so universally important a consideration in mental hygiene as is sometimes believed.

Here it may be mentioned that the methods so far used in attacking the problem of diseases suffer from certain limitations, which will be considered in detail in Chapter III; and that, therefore, the conclusions reached one way or another ought to be evaluated with those limitations in view.

4. *Mortality and Longevity.* In his *Statistics of Families* Ansell (4) has a chapter on the rate of mortality among children classified according to the numerical order of their births. In estimating the order, no distinction was made between those born alive and those born dead; and multiple births were considered

as single ones. The total number of children studied was 48,843.

One of the features reported was the large number of stillborn eldest children, the proportion of dead to living births in that class being more than double the average of all the other classes. In the first week of life the mortality among eldest children was also much greater than in other classes. But after that period it fell below average for several years, and at most ages, up to the end of the sixth year, the eldest children showed the most favorable rate of mortality of all.

The mortality curve of eldest children differs from that of younger children also in two other respects, viz., first, in the progressive diminution in the death rate during childhood terminating earlier and changing less abruptly to an increase; and secondly, in there being another well-marked diminution after the twenty-second year.

It is also shown in Ansell's life-tables that, out of each 1000 children born alive, 846 of the first born survived to age fifteen, 848 of the second born, 847 of the third, 838 of the fourth, fifth, and sixth together, and 813 of the seventh and following children. For all children together, the figure was 844. This comparison seems to indicate that it was the second and third children who had the lowest rate of mortality. Ansell's statistics referring to the numbers of 1000 living born children who were alive after 45 years are the following: first-born 658, second 653, third 655, fourth to sixth 654, and seventh and later 635. It was pointed out by Ploetz (1909) that the relatively small difference of the various birth numbers in Ansell's material might be ex

plained by the fact that it consisted of well-to-do families.

Geissler (50), on the basis of 26,000 legitimate births in families of miners, found that infant mortality rapidly increased after the fourth child. The figures are presented in Table 6.

TABLE 6

Order of birth	Infant mortality per 100 births	Order of birth	Infant mortality per 100 births
1	23	7	31
2	20	8	33
3	21	9	36
4	23	10	41
5	26	11	51
6	29	12	60

Westergaard (132, pp. 190-191) took Ansell's figures and calculated the probability of a person's dying in an age interval before reaching the text. The results are presented in Table 7.

In this table the higher ages do not show anything of

TABLE 7

THE NUMBERS IN PERMILLES OF THE PERSONS IN EACH AGE INTERVAL WHO WOULD DIE BEFORE REACHING THE NEXT AGE LIMIT

Age in years	Birth order 1	2	3	4	5
0	67	49	47	54	70
½	21	22	23	26	30
1	20	21	21	25	29
2	26	29	25	21	31
5	24	25	29	31	28
10	23	20	19	20	15
15	70	67	72	69	72
25	76	83	83	81	78
35 (Next 45)	84	98	90	86	87

importance; but, with younger ages, a significant phenomenon is apparent. The first-born as well as the last-born have a high mortality in the beginning, but in the second half year the mortality among the former becomes lower than in the other groups. With some fluctuations, this tendency continues up to the fifth year. The later-born appear to be comparatively weak.

Bell (11), in a similar study, reported that the majority of the persons who were the only children of their parents died young (58.5%), but there were only 41 cases.

Pearson (97) found that stillbirths were most frequent among the first-born, that infant mortality was higher among the first-born, and that the health of the first-born child, during its early years, was below par.

Bolt (15) reported in Abt's *Pediatrics* that "on the average the first-born children have a 33% higher mortality than the second-born. The still-birth rates for males and also for first-born infants are in excess of those for females and for second and following births."

Redfield (103) selected from the Redfield genealogy every family which had four or more sons who had reached maturity and who had not lost their lives by war or by accident. The average life of different sons was found to be as follows: eldest son, 60.85; second son, 69.14; third son, 69.85; fourth son, 71.14. This result shows an increase in length of life as birth ranks grow higher.

In apparent contradiction to the above findings, Beeton and Pearson (10), studying a thousand or more pairs of siblings, found their average age at death to be as follows:

	All adult	Elder adult	Younger adult
Sister	57.795	59.924	55.667
Brother	56.568	58.560	54.575

This result appeared so interesting to them that they secured 1051 pairs of brothers and 733 pairs of sisters where the interval between births was known and further studied them. The data are not quite the same as for the pairs of adult brothers and sisters given above, but they show much the same advantage, i.e., four years to the elder. They have further demonstrated that longevity is correlated with position in the family, and their conclusion is that, on the whole, the earlier-born members are fitted to survive the longer.

Ploetz (99-101), in a study of 3300 children of royal families, found that the rule of increasing mortality does not hold true for these children where mortality remains constant almost up to the ninth child. Upon the basis of such findings, the author infers that the increase of mortality with family size is due to the fact that large families make it more difficult to secure the necessary care and nourishment for the children.

In reporting an inquiry which covered about 300,000 government employees, March (85) concluded that the mortality increased with the order of birth. This was the case irrespective of sex and of the status of the parents, and applied to both children and adults. The age of the father was found to have no marked influence on the mortality of the children. This increased mortality according to the order of birth was explained by the fact that, in one and the same family, the last-born did not appear to be subject to greater mortality

risk than the first but that the mortality of children of large families was greater than that of small families.

Popenoe (102) made a similar study. His data were obtained from the Genealogical Record Office of Washington, D. C. The individuals were, in general, over 90 years of age, some living and some dead. A few were included whose age was less than 90, but more than 80; these, however, represented a very small proportion of the total. The results, taking only the families of two or more children, and omitting decimals in the second column, are as shown in Table 8.

TABLE 8

Distribution of the Aged According to Birth Order

Birth order	Expected	Found	Birth order	Expected	Found
1	132	201	9	26	22
2	132	118	10	16	20
3	121	104	11	9	4
4	101	95	12	5	8
5	82	82	13	3	4
6	65	40	14	1	2
7	50	53	15	0	3
8	38	30			

On the basis of these results, the author concludes: "The first thing to strike the attention is that the first-born, who have sometimes been supposed to be handicapped with all sorts of physical weaknesses from birth, are relatively more numerous than any other birth rank. But the number of individuals in the last birth ranks is somewhat surprising, in view of the generally accepted belief that the last child of a very large family tends, because of uterine exhaustion in the mother, to be defective, and is often an imbecile of the so-called Mongolian type. The table here given shows at least

that the last born of a large family is not necessarily lacking in a tough constitution which will enable him to survive for ninety years or more" (p. 396).

Reiter (104) reported that his findings agreed with those of Dresel and Fries in showing the increase of mortality with "Gebürtigkeit," but disagreed with the latter in failing to indicate the counteracting influence of the economic factor.

Holmes and Wilson (72) made a study of the relationship between the order of birth and the longevity of the child. Their source of data was Allström's *Dictionary of Royal Lineage*. A selection was made of all those royal families in which four or more members lived to be 21 or more years of age and died a natural death. The results are shown in Table 9.

The authors have concluded that their findings are not in agreement with either those of Pearson or those

TABLE 9
LONGEVITY ACCORDING TO BIRTH RANK

Birth rank	No. of cases	Av. age at death
1	264	51.6
2	257	52.4
3	266	50.6
4	253	54.2
5	214	51.9
6	173	53.4
7	115	54.1
8	88	50.5
9	77	47.0
10	50	55.2
11	30	50.5
12	26	56.4
13	15	56.8
14	11	48.7
15	7	53.1
16	5	51.8
17	2	41.0
18	2	47.0

of Redfield, for they do not indicate any relationship between the order of birth and the expectation of life. They also have pointed out the fact that their "data are based on royal families while those of Pearson and Redfield deal with ordinary humanity . . . royalty has several advantages over the rank and file of people in general, on account of the fact that there would be little correlation between lateness of marriage and status or between size of family and those extraneous influences which cause the size of family to vary so greatly in the general population. While royal families may be voluntarily limited like other families, the inequalities of the limitation are probably not so great as among people of diverse social and educational advantage."

In a more extensive study which includes data on all the individuals concerning whom the ages at death could be obtained, Holmes (71) has found that "within each group of families of a given size there is little indication that the members of any birth rank live longer than those of any other." In order to ascertain what relationships were indicated by the data as a whole, order of birth was correlated with longevity of children, and a negative correlation of $-0.090 \pm .012$ was found. "This means that children of higher birth ranks do not live as long on the average as those of low birth ranks. The correlation is a weak one, to be sure, but it is over seven times its probable error and is therefore not due to mere chance."

The tendency for the earlier-born children of a family to live longer than the later-born was tested in another way by comparing those pairs of siblings whose births were separated by the greatest number of years.

It was found that the average age of the elder sisters living over 20 was 52.737 and of the younger sisters 52.958. The average age of the elder brothers was 52.105 and of the younger brothers 50.375.

Looked at on the surface, the results are in utter disagreement; but in fact, the high mortality of the first-born at birth or during infancy does not necessarily mean their high mortality in later years, and, on the contrary, there is every reason to believe that a more stringent selection at birth will result in lower or perhaps equal mortality in adulthood. The latter assumption seems to be to a certain extent confirmed by studies of Ansell and Westergaard.

5. *Fertility and Sterility*. To this aspect of the problem, only a small amount of attention has been given. There are two studies which may be worth a brief mention here.

Ewart (42) has studied 1110 sons. According to his findings, the correlation between the order of male birth and the number of children born to son is $.021 \pm .020$, and the correlation between the order of male birth and the number of children born to son with age at marriage and death of son or present age constant is $-.037 \pm .019$.

Jones (77) has reported a study of five genealogies. Each genealogy furnished approximately 4000 individuals, and the five together furnished a total of 20,412 individuals. The 20,412 individuals were distributed throughout approximately 4000 such household families, each of the genealogies holding about 800 of this total figure. In Table 10 the average fertility of the total population stock included in the genealogies and

that of the first-born are compared both for the 50-year periods and for the entire period from 1651 to 1900. In Table 11 the percentage of the total number of children born who married is compared with the percentage of the first-born children who married. These two tables show that the percentages of the total stock and of the first-born who married were the same, and that, of those who did marry, the fertility of the first-born and that of the total stock were practically the same.

B. Behavior Traits

1. *Intelligence.* Willis (133) tested almost two thousand children by the Stanford Revision of the Binet

TABLE 10

Fecundity of First-Born Children

Period	Average fertility of total stock	Average fertility of first-born
1651-1700	5.8	6.4
1701-1750	5.9	5.8
1751-1800	6.1	5.7
1801-1850	4.8	4.9
1851-1900	3.0	2.8
Total	4.2	4.0

TABLE 11

Percentage of Population and of First-Borns Married

Period	Percentage married of total stock	Percentage married of first-born
1651-1700	.76	.77
1701-1750	.72	.71
1751-1800	.74	.73
1801-1850	.69	.69
1851-1900	.62	.60
Total	.69	.60

Scale in the Alex Taylor School, Edmonton, Canada. There were no low-grade defectives in the school population. Two hundred nineteen pairs of children were obtained, each pair being, respectively, the first and second children in a family. The median difference in favor of the second child was +4.5. The sigma of this difference was 1.16. Therefore it was almost certain (9997 chances out of 10,000) that the median difference in IQ between the first and second children lay between +1.02 and +7.98. This investigator was of the opinion that the higher incidence of mental deficiency among first-born children might be due in part to the fact that they were the first-born children in families in which the rest of the children were properly classified only as "border-line" or "backward."

Sutherland and Thomson (119) tested 1084 children with ages ranging from 10-6 to 11-6. With the size of family kept constant, correlations were computed between IQ and birth position for families of 4, 5, 6 and 8 children. None of the coefficients differed reliably from zero.

Arthur (5) made a study of intelligence in relation to position in the family. A large percentage of her subjects were of foreign birth or parentage, and most of them entered kindergarten with little or no knowledge of the English language. They were tested with the Kuhlmann 1917 Revision of the Binet Scale. According to the results for 92 pairs of siblings with Finn, Russian, or South European surnames, all of whom had been tested with the Kuhlmann-Binet scale at the end of one year in the kindergarten, the average IQ of the elder sibs was 93.05 with an *S.D.* of 11.29, and that of

the younger was 99.14 with an *S.D.* of 10.42. The difference was in favor of the younger sibs and was found to be reliable. To determine the importance of the language factor, the study was extended to 271 pairs of siblings of kindergarten and grade-school age. These included the 92 pairs described above, and also 179 pairs the elder of which had the advantage over the younger of one year or more than one year of additional training in English. In this group, the average IQ for the elder sib was 89.3 and that for the younger was 96.9. This difference in favor of the younger sibs was also a reliable one. According to the author, the phenomenon of decreasing IQ with priority of birth was further emphasized by the data from families of which three or four children were given Kuhlmann-Binet examinations, although it was stated that these families were selected largely upon the basis of school difficulties on the part of the elder children. However, from a study of 70 pairs of sibs in American families each of whom had been given a Kuhlmann-Binet test between the chronological ages of 5-0 and 7-0, it was found that the average IQ for the elder sibs was 105.9 with an *S.D.* of 11.23 and that for the younger was 106.9 with an *S.D.* of 10.67. The difference was slight and unreliable. Upon the basis of these findings, the author was inclined to think that the birth order differences that were found in this study were attributable to factors of environment rather than to those of nativity.

Commins (28) compared the IQ's of 142 pairs of siblings chosen from school grades 3 to 8. These IQ's were obtained by means of the McCall Multi-Mental Test. In 99 cases, the younger member of the pair had

the higher IQ; in 43 cases, the older, the average difference between the sibs being 17 and 10.3 points, respectively, in the two groups. The interpretation was offered that the McCall test was probably so constructed that the IQ tended to decrease with age. However, the author was inclined to feel that, in the light of the agreement of this result with those of Miss Arthur, the thesis that the younger children in a family are the brighter need not be abandoned.

Jones (78), with the assistance of the present reviewer, made a study of 614 pairs of juvenile and adult siblings. The test scores used were those of the Stanford-Binet for children up to the age of 14 and those of the Army Alpha for children from 10 upwards. Each mental age and Alpha point score was translated into a sigma index, dividing the deviation from the age median by the *S.D.* for that age. In this group of comparable scores, the pair differences were computed for pairs adjacent in birth order and for non-adjacent pairs.

TABLE 12

No. pairs	Adjacent pairs 330	Non-adjacent 284	Total 614
Mean difference (Older-younger)	—.010	—.055	—.030
Median difference	+.035	—.058	—.003
$S.D._{dist.}$	1.085	1.070	1.075

Table 12 shows that there is no difference in either central tendency or variability when the older and younger sibs are compared. Pairs successive in birth order are no more similar than pairs separated by one or more intermediate sibs.

Thurstone and Jenkins (121) have recently made a

study of the relationship between birth order and intelligence. Their material was drawn from the records at the Institute for Juvenile Research. There were 382 families for whom the records of both the first- and second-born children were available. The comparison was confined to siblings, who ranged from the first- to the eight-born. The mean IQ of the first-born was 81.75, while that of the second was 84.84. The comparison therefore shows a slight advantage for the second-born. It was also found that 23 out of 28 comparisons turned out in favor of the later-born. The authors claimed that the possibility of defects in the standardization of the Stanford-Binet tests for the successive ages was ruled out in two ways. They had analyzed the test records of the Institute separately for each age, and had tabulated their records in terms of standard scores for each age group, but the final results remained the same. A further check they mentioned was a study in progress by Miss Steckel. She had given four group tests to the public school population in a city, and the records of all siblings had been compared. Since that study involved the testing of a large school population, the tests were standardized on the experimental population itself. The published norms were not used. The test performances were reduced to standard scores for each age, and the comparisons were limited again to the siblings. The preliminary findings in that study were in agreement with the results of Thurstone and Jenkins' investigation.

From the above, one will note that, in the studies which indicate the existence of a difference in IQ between the first and subsequent children of the family,

certain aspects of the statistics have failed to receive due consideration. These will be discussed in Chapter III. As to the other studies of intelligence, either the findings are inconclusive or the primary interest does not lie in the status of the first-born.

2. *Talent and Subnormality.* Galton (47, pp. 33-34) made a study of English men of science. He found that, in percentages, their positions among their brothers and sisters were as follows: only sons, 22 cases; eldest sons, 26 cases; youngest sons, 15 cases. Of those who were neither eldest nor youngest, 14 came in the elder half of the family; 12 in the younger half; and 11 in the middle. These results purport to show that elder sons appeared nearly twice as often as younger sons, that, as regards intermediate children, the elder and younger halves of the family contributed equally; and that only sons were as common as eldest sons. Here it may not be out of place to mention his study (48, p. 78) of 72 English judges. The eldest sons have not made half as many careers as judges as the younger sons, but he regards social influence as responsible for the phenomenon.

Yoder (135) examined the early lives of 50 eminent men of various nationalities belonging to the eighteenth and nineteenth centuries. It was found that the youngest sons occurred oftener than intermediate sons and eldest sons oftener than youngest.

Cattell's (25) study of American men of science also yielded interesting results, which are presented in Table 13.

Gini (53, 54) claimed that the first-born predominated among the professors in Italian universities. His

TABLE 13
DISTRIBUTION OF AMERICAN MEN OF SCIENCE ACCORDING TO BIRTH ORDER

Family size	Birth order			
	First	Second	Third	Fourth
2	57.4%	42.6%		
3	44.0%	31.2%	24.8%	
4	36.1%	22.4%	21.8%	19.7%

method of investigation was to send questionnaires to the professors. Four hundred forty-five replies were received, and 416 of them related to families of two or more. His statistical procedure, however, was not described.

In Ellis' *A Study of British Genius* (40) 1030 people (975 men and 55 women) were included. They were those persons to whom three pages or more of space was allowed in the *Dictionary of National Biography.* One of the problems studied was "In families of different sizes, what relation do eldest genius children bear to genius children of intermediate position?" Results of this inquiry are reproduced in Table 14.

TABLE 14

Family size	Position of eminent child		
	Eldest	Intermediate	Youngest
2	15	0	12
3	15	6	11
4	10	16	3
5	10	18	7
6	8	20	6
7	15	14	5
8	2	17	4
9	8	7	4
10	5	10	3
11	3	12	2
12	1	10	2
13	1	4	2
14	0	5	2
Over 14	1	9	4

The author concludes: "It would appear that there is a special liability for eldest and youngest children to be born with intellectual aptitudes, the liability being greater in the case of the eldest than of the youngest, for there are altogether 94 eldest children to 67 youngest children, the intermediate children numbering 148; or 30 per cent are eldest children, 21 per cent youngest, and 47 per cent intermediate. It will be seen that while the eldest and youngest children of ability absolutely outnumber those of intermediate position, notwithstanding the large average size of the families producing children of ability, and the consequently much greater number of chances possessed by the intermediate children as a group, the chances of the eldest attaining eminence as compared with the chances of the youngest are not the same throughout. In the small and medium-sized families it is the eldest who most frequently achieves fame, in the larger families it is the youngest. It may be added that if we were to take into the consideration the survivors of a family only (the net fertility) the youngest children would occupy a still more conspicuous position" (40, pp. 102-103).

Among 754 gifted children, Terman (120, pp. 121-125) found the following percentages of first-born: 56.1 in families of two, 36.9 in families of three, and 33.0 in families of four. He pointed out that the fact that the completed families in his group showed an atypically small proportion of gifted first-born was for the reason that his main search was confined chiefly to grades 3 to 8 and hence yielded few children above 13 years of age. The eldest child in a completed family would frequently, if not in a majority of cases, be

older than 13, and would therefore be less likely to be caught in his survey. "Possibly in the case of incomplete families, those with gifted first-born were more likely to be caught."

All the above investigations seem to indicate that the distinguished persons are likely to be the eldest children of the family, but Mlle. Robinovitch has been quoted by Tredgold (123) as having brought forward some evidence to show that distinguished men are more frequently the last-born.

A similar study was reported by Key (81). Her problem was to determine if there was any scientific basis for the popular notion that the eldest child has the best chance of obtaining prominence. She studied the family records of many famous Americans, from Alexander Hamilton to Mark Twain, but her conclusion was that the order of birth in a family had nothing to do with their achieving fame.

The several studies given above deal with the problem of superiority. Now let us turn our attention to the problem of inferiority. According to both Pearson and Ellis, A. Mitchell (88) was the first one to point out many years ago that, among idiots, the youngest born and especially the eldest born largely predominate over the intermediate children. Among 443 idiots and imbeciles, he found 138 to be first-born, or 31.1%. Excluding a large contingent of illegitimate children, who are, of course, disproportionately often first-born, and confining his observations to cases in which only one idiot, and that a congenital one, occurred in a family, he concluded again that "idiocy is more likely to occur among first and latest pregnancies than among

others," since of all the children born in Edinburgh and Glasgow in 1855, 22.8% were first pregnancies, while of the 85 idiots 33% were first pregnancies.

Down (34) claims liability to idiocy among primiparous children because 24% of the 2000 idiots (illegitimates largely excluded) examined by him were first-born children. But, in view of the large size of the families represented, the first-born children were probably born of very young mothers. Such being the case, the heavy incidence of idiocy among the first-born becomes a question of mother's age and not one of birth order. Moreover, the factors that conditioned the examination may have seriously biased the sample, and the criteria for idiocy used in those years may have been too crude to be trustworthy. It may also be mentioned that he found that 40% of idiots born in a state of suspended animation were primiparous children, and that he invoked, as a cause of this as well as the above phenomenon, the "exalted emotional life" of newly married couples and the increased difficulties of a first parturition.

Shuttleworth (114) studied the children at the Royal Albert Asylum. Five per cent of them were Mongolian imbeciles. Some 40% of such children were the youngest members of large families. His explanation was that anything that depressed the procreative powers of the mother might be a factor in the production of the cases. He had commonly found a family history of phthisis, and had observed that Mongols themselves, when they survived infancy, were prone to phthisis, and that those who died young succumbed in many instances to some form of tubercle.

Rivers (108) has quoted Hunter as reporting that, among idiots and their parents, there was heavy incidence at either end of the family. Wallich and Fruhinholz have been referred to by the same author (97) as stating that 16 out of 30 premature children brought to a Bicêtre clinic for nervous affections and other defects were eldest born. With regard to the proportion of the first-born among mental defectives, Duncan has been quoted by Rivers as giving the following summary: Among Mitchell's 443 idiots and imbeciles 138 were first-born; among Wilbur's 675 there were 191 first-born; among 100 of Beach's, 20; among 2000 of Down's, 480. In other words, among 3218, 829 or about 26% were first-born, and presumptively born of young mothers.

In a study of 994 feebleminded cases admitted into an Anstalt für Geistesschwache for a period of about 20 years, representing as many marriages with a total number of 5915 children, S. Hansen (61) finds that, instead of the expected number of 167 for the first-born, there were actually 234 of them, but that no special difference existed for the succeeding birth numbers. For the second-born the expected number and the actual number are 161 and 159, respectively, and for the third they are 148 and 149. The material obtained from the Kellerschen Anstalt also shows that it is the first-born that are most frequently found.

It should be noted that this author has mentioned the following sources of error in his material: First, families are incomplete in many cases. Secondly, there is a continual decrease of fertility over a long period of time. Thirdly, the material is very mixed; it repre-

sents various social strata with different degrees of fertility. In addition to these, Pearson (97) has pointed out a possible source of error, namely, that the material recorded is spread over a score of years, so it is quite possible that the same family is several times represented.

However, Pearson's (97) own investigation also shows a considerable excess of first-born above what was to be expected. There is no preponderance of second-born defectives; when Mongolians are excluded, the mentally deficient are fewer than expected up to the eighth-born, and show a slight excess in large families of eight and over.

In discussing the possible causes of feeblemindedness, Tredgold (123) has touched upon the question of primogeniture. We may quote from him as follows: "Certainly my own experience is to the effect that it is more common for the later-born and not the first-born to be affected. In those families in which there is a pronounced tendency to mental and physical degeneracy, the effect usually appears to be more and more marked upon each successive child, and often enough the idiot is actually the last born. I have notes of not a few families in which the first one or two children presented no great departure from the normal; these were followed by one or two others who succumbed to ordinary children's ailments from which healthy children would probably have recovered; then came the idiot, in some cases to be succeeded by a number of stillbirths."

Kuhlmann has been quoted by Arthur (5) as having compiled a table from the data of Piper, Koenig, and

Kerlin to show the incidence of feeblemindedness in the family group. According to the results, the incidence of feeblemindedness is the highest among the first-born and decreases for succeeding positions.

Ordahl (93) studied the birth rank of Mongolians, and found that a disproportionate number of Mongolians occurred among the later-born of large families as compared with the earlier children of the same. parents. This agrees with the findings of both Shuttleworth and Pearson.

In a comprehensive study of Mongolism by Brousseau (19) this question of birth order has been considered. She has quoted Spicer as having observed, on the basis of 23 cases, that the Mongols came often at the end of a large family, but several of the cases were also the first children of young mothers. Hill has also been quoted as having found, in an analysis of 8 cases, that fully 50% of Mongolian idiots were later members of a large family. However, in Brousseau's own material, it is noted that the Mongol is frequently the last child of the family, although not necessarily of a large family. In 99 cases out of 583, or 16.98%, the Mongol was the only child; in 336, or 57.6% the Mongol was the last child of a family having at least two children. The author has pointed out the following possibilities for consideration: In the first place, there may have been a few cases in which other children were born some years after the Mongol was reported as being the last child. In the second place, as birth control is practiced very frequently, it is reasonable to suppose that intelligent parents may, after the birth of a defective child, have decided to have no more children,

fearing that later offspring might also prove mentally deficient. In the absence of statistics regarding these facts, the author hesitated to make conclusions concerning the causal relation between Monogolism and birth position.

Recently, Dayton (31) has reported a study of 10,410 retarded children examined by the Stanford-Binet. He considered families of different sizes separately, and found no evidence that the feeble-minded child tended to be either the first or the last in the family.

We may also briefly state the opinion of the following men on the intelligence of the first-born. According to Voisin (108), among the children of an alcoholic father the first is often an idiot, and the second only an imbecile. Shandy is quoted by Ellis (40) as asserting that the eldest son is the blockhead of the family. But Campbell (22*a*), in his *Causation of Disease,* combats the belief that the first-born child is inferior.

These findings seem to indicate that there are arguments for and against the first-born. However, it should be borne in mind that, apart from the other considerations, the methods used in most of these studies were far from adequate; and therefore, that the preponderance of the first-born merely among the super- or sub-normals can hardly be regarded as convincing evidence. In this connection, it may also be mentioned that, according to Gini (53), the first-born are more variable than their sibs; and that, according to Weinberg, the first-born die young less frequently and, therefore, have more opportunity to realize their various tendencies than the later-born; and that, according to Ploetz (99), the first-born children studied

as his statistics concern only the deaths in the first year, it remains a question whether or not the same conditions will prevail in the period in which criminal tendencies are supposed to develop. Goring's data were afterwards analyzed by Pearson, whose conclusion was that the actually observed first- and second-born criminals were 717 as against 557 which would have been anticipated if the tendency to crime had been divided equally among all members. There was a deficiency of both intermediates and last-born criminals.

A study of juvenile theft was reported by Baker and Decker (7). The experimental group and the control group each contained 84 boys. The experimental cases had been formally convicted of theft by a juvenile court; and, through that organization, the confidential records were made available for the study. The control group was selected by the schools, and the cases were matched as to age, grade, nationality, and neighborhood. The median age was 13 years, and the median grade was the upper sixth. It was found that the position of boys among siblings was not a significant factor.

In view of the conflicting evidence concerning the criminality of the first-born and in view of the inadequacy of the methods used, we should suspend generalizations until further evidence is available.

4. *Other Traits of Behavior.* In 1899, after examining 1507 Michigan school children, Carman (23) concluded that first-born boys were more sensitive to pain, as estimated by the temple algometer, than second or subsequent children. In general, sensitiveness to pain was found to decrease in order of birth, the exception being that later-born girls were slightly more

sensitive on the right temple than were the second-born. Carman thought that if the number of second-born girls were larger this exception might not occur. Most of the children were from laboring classes, by which is meant artisans and unskilled laborers.

These findings are in agreement with those of McDonald (87), who stated that of 53 university students the first-born were more sensitive to pain than the second-born.

In Wells' (131) study referred to above, it was found that, among a large number of school children at San José, California, whose positions were noted as regards weight discrimination, reaction-time, voluntary action, ability, endurance, mental ability, neatness, and deportment, the eldest tended to show best, and the youngest children, while inferior to eldest, were superior to intermediate children.

Reynolds (107) studied the first 492 applications to the habit clinics of the Massachusetts Division of Mental Hygiene in the first 18 months of their operation. Sixty-seven and five tenths per cent were under 6 years, 84% were under 8 years, and all but 2.5% were under 12 years of age. "The position of the child in the home seemed to have some, though not marked, significance; 47 were only children, 80 had no rival children of their own sex in the family; 108 were the oldest, 83 were the youngest, 24 were adopted or placed-out children; 26 were cursed by being the favorite child, usually of one parent, in 3 cases of both; 85 were regarded as peculiar by their families and usually advertised as such among their parents' friends."

Goodenough and Leahy (56), from the files of

Demonstration Child Guidance Clinic 2, which was conducted in the Twin Cities during 1923-1924, obtained a total of 322 consecutive Minneapolis cases, and studied them with the view to determine the proportion falling within each of four general groups considered, viz., oldest children in families of more than one, youngest children in families of more than one, only children, and children occupying intermediate positions in families of three or more, who for convenience will be referred to as "middle" children. As to the principle followed in those cases in which the death of a sibling involved a change in the group classification, e.g., when the death of the first-born child left the second-born as the oldest living child, it was decided to ignore all deaths occurring before the clinic patient was three years of age, and to include as if living all dead siblings whose death occurred after the clinic patient had reached the age of three. A small number of adopted children were included. Classification of the 322 cases according to this plan brought the following results.

	No. of cases	Percentage of total
Oldest	98	30.4
Middle	102	31.7
Youngest	81	25.2
Only	40	12.7

Here it is evident that what was considered was the socially first child and not the biologically first one. Moreover, there is no statement as to whether or not these families were complete.

They also made a study of 293 kindergarten children at the Minneapolis Child Guidance Clinic. The children were rated by their teachers upon 14 traits. On

the basis of their records, they were grouped into the classes: oldest, middle, youngest, and only children of their respective families. The oldest children showed significant tendencies toward lack of aggressiveness and self-confidence, lack of leadership qualities, much suggestibility and some seclusiveness and introversion. The middle children showed these traits to a lesser degree. The youngest showed individual variation. The only children showed higher ratings for aggressiveness and self-confidence, for gregarious interests, and for instability of mood and flightiness of attention.

Stratton (118), in a study of anger and fear, has touched upon the question of primogeniture. His general procedure was to have students of psychology classes grade their own behavior according to a prescribed scale in certain situations supposed to provoke fear or anger. No difference was found between the first-born and others in regard to fear, but a difference existed in respect to anger. The numbers of the first-born and later-born were 302 and 544, respectively, in the case of anger and 273 and 441, respectively, in the case of fear. When the first-born were divided into two classes, of those who had siblings and of those who had none, the number in the latter group was too small to insure the reliability of results. However, so far as the evidence goes, the fact that the first-born are apt to be more irascible does not seem to be due to the presence or absence of siblings.

In Busemann's (21) study of the group of approximately 400 children in attendance at the Griefswald Mittelschule, which was chosen as representing socially and economically a sufficiently uniform and average

group so that valid comparisons of families of varying size and composition could be made within it, sibling status was compared with four behavior manifestations, on the basis of simple averages, namely, (1) school status; (2) school marks by groups of school subjects; (3) degree of reflection measured (*a*) by ratings on extended answers in writing to the question, "Are you satisfied with yourself?" and (*b*) by pupils' ratings on the "saddest" and the "most cheerful" pupil; (4) teachers' ratings on "docility" and "hyperactivity." The results show that, in general, children are more industrious and able at school the more siblings they have up to four; that lack or fewness of siblings has a strong influence, favoring reflectiveness (introversion), feelings of self-dissatisfaction, hyperactivity and unrest, and that this condition is inimical to school success; that the eldest sibling has on the average a higher class position than the median sibling, the youngest a lower position than the median; and that having siblings of the opposite sex is disadvantageous for school success, but has no effect on "reflectiveness." Poverty of siblings apparently favors variations, for the most part in the direction of less school accomplishments and unfavorable traits, but occasionally (in the case of girls) in a favorable direction. In short, the sibling status affords an operative influence in every case.

Goodenough (55) reported a study of 1897 observations of the emotional behavior of 990 children in a mental-test situation. The ages of these children ranged between 18 months and 6 years. Ratings on shyness showed that only children were less shy than the youngest children, but the oldest children were not

reliably different from the others. Only children also appeared to be somewhat more distractible than those in the other two groups. But the ratings on negativism showed no differences, related to position in family, which were not within the limits of chance.

According to Smith (115) there is no conclusive evidence as yet to substantiate the contention that the younger children of a family have an advantage over the older ones in learning to talk. In the published data on two children of the same sex in the same family studied at the same age, the younger child appears to have the advantage in the cases reported by Schneider (113), Stern (116), Holden (69), Gale (46), and Bateman (8); but the older child is the farther advanced in the cases reported by Block (13), Gheorgov (51, 52), and West reported by Gale (46*a*), and Nice (91, 92).

In studying the relationship between onset of speech and intelligence, Abt (1) and others have considered the possibility of differences in favor of the younger members of the family. Their data were drawn from the records of the Institute of Juvenile Research in Chicago, consisting of 1000 cases, 500 boys and 500 girls. They are studies of children referred to the clinic for behavior difficulties as educational problems, for habit training, or for advice as to institutional and other placement. The average IQ is 80 for the boys and 82 for the girls. The range in the age of onset of talking extends from 6 to 60 months for both sexes. The size of family varies from that of one to that of thirteen. There was found absolutely no relationship between the ordinal position of a child among his sib-

lings and the acquisition of speech, the coefficients of correlation being 0.03±0.03 for the boys and 0.00±0.03 for the girls.

In addition to the above studies of the miscellaneous phases of behavior, we may also mention the following observations:

Hug-Hellmuth (74), in his article "Vom 'mittleren' Kinde," says: "It is the middle of three children of the same sex differing little in age which feels most painfully the uncertainty of its position at home. The unconscious childish discord becomes especially dangerous to the second of three sisters if not only the oldest but also the youngest has had a happy choice of love; the marriage of the youngest not seldom signifies the middle one's failure in love affairs. Much more favorable is the fate of the girl between two brothers. Under the influence of the elder, the qualities of 'kind-Weib' will be unfolded; and, to the younger, she will owe a strong motherhood. The dark side of a middle child's life is spared, most of all perhaps, of the boy between two sisters. If he is pretty close to the eldest sister in age, the first place in the nursery will be given to him."

Adler (3) has described the characteristics of the first, second, and third child with special emphasis on the psychological mechanisms involved. We may quote him as follows: "The oldest child feels dethroned by the coming of his brother and wants to restore his place by fighting. Unless he can overcome in the struggle for supremacy in his universe he is apt to become depressed, peevish, more or less hopeless, and will show his hopelessness later in life if confronted by problems. He is very likely to be conservative, to understand

power and to agree with it. If he is strong enough he becomes a fighting child.

"As for the second child he is never alone, but is always confronted by the older child. This constant picture before him of an older and bigger child begets in him a sense of rivalry. If successful, he is an excellent type, but if defeated, for instance, if he is not able to compete successfully with the older child in work and in play, he loses hope, becomes depressed and has a bad time of it.

"The third child has to fight for a place in the sun, but he has no successor. This gives him a great sense of power, and if he is capable, he often overcomes the older children in the family by his sense of importance. If he is not capable, he perhaps hides behind the fact of being spoiled, and becomes lazy, escaping from task, wasting time and making excuses."

C. Other Phases of the Problem

From a study of 16,301 families, Duncan (38) drew the following conclusions: First, the majority of earlier-born children up to the third or fourth came into the world in more rapid succession than those that immediately followed. Secondly, the majority of children numbering from the fourth or fifth on to the tenth succeeded one another more slowly than those that came before or after. Thirdly, the majority of children following the tenth came hurrying after one another with a gradually increasing rapidity, which exceeded that of all their predecessors.

Burdge (20) made a study of 17,627 employed boys in Greater New York. He found that only among

boys of foreign parentage the first-born tended to leave school earlier, but he considered that the greater size of families among the foreign population was responsible for this fact.

The relation between the order of birth and the duration of nursing was studied by Schlesinger (111). He investigated 1445 large families in Frankfurt. The total number of children was 8078, with an average family size of 5.6. His results are shown in Table 15.

TABLE 15

THE DURATION OF NURSING

(Figures in percentages of the total number of mothers)

Birth order	No. of cases	Nursed not at all or at most a week	Months 1-1½	2-3	4-6	7-9	10-12	13 and longer
1	398	36	13	22	12	7	7	3
4-5	294	21	9	16	17	13	13	11
6-7	278	27	9	16	13	11	12	12
8-9	82	22	8	13	19	7	14	18
10-12	92	23	5	10	17	11	22	12
13 and later	23	21	4	13	17	17	17	9

Here it is shown (pp. 58-59) that "scarcely two-thirds of the first-born participated in the blessings of the mother's breast, while three-fourths or four-fifths of high birth numbers are more or less nursed. Still greater and more significant is the difference with respect to the duration of nursing. The first-born, on the average, received the breast for only two months; the eighth-born, for over a half year. The later-born, the eighth to the twelfth, show the most favorable figures; over a third of them were nursed for ten months and longer, whereas only 10 per cent of the first-born enjoyed the breast so long. The decrease of nursing

capacity of the mother with the increasing birth number is scarcely recognizable in my table; the 13th children were nursed scarcely less than the 10th. . . . Up to the fourth child an increase of nursing duration is the rule; the fifth and later-born children are nursed seldom longer than their preceding ones. However, the last of a series of children is very frequently nursed only for a short time."

General Summary

By way of summary, I wish to call attention to the following facts: First, at birth the earlier-born seem to be inferior in height and weight to the later-born, but surpass them soon afterwards in these respects. Secondly, with regard to diseases, there seems to be need for an improvement of methods of attack. Thirdly, apart from economic considerations, the problem of the mortality of various birth orders appears to be reducible to the problem of family size; in other words, differences found in the mortality of various birth ranks may be due to the different rates of mortality in families of various sizes. The preponderance of stillbirths among the first-born, however, may have other causes. Fourthly, no difference in fertility has been observed between the first- and later-born. Fifthly, studies of intelligence and of talent and subnormality appear to be far from satisfactory, and, owing to the important bearings of such studies upon human life in general, further research is imperative. Sixthly, certain studies of emotion as well as personal observations have opened new vistas for experimentation. Finally, the remainder of these studies may serve as stimuli to further efforts.

III

A STUDY OF INTELLIGENCE AS A FUNCTION OF BIRTH ORDER

The Problem of the Present Research

The problem of the present research concerns the relationship between the position of birth and the intelligence of the child. Limitation of the study to the first two birth orders has been deemed desirable because of considerations of sampling.

In the following pages, the term "intelligence" refers to test scores (converted into I.Q.'s) of the Stanford Revision of the Binet Scale, the Terman Group Test of Mental Ability and the National Intelligence Test. To this definition of intelligence, the objection may be raised that different tests may measure different processes, and therefore that the scores obtained on them can hardly be taken in one sense. But, as a matter of fact, we are under no necessity of taking the scores obtained on different tests as coordinate or identical in nature, since the siblings are to be compared on the same test and not on different tests. A further possible objection is that, even in one and the same test, the processes measured at the different age levels may differ in nature and therefore the comparability of the scores obtained at the different age levels is questionable. This objection is naturally meant to apply to the Stanford-Binet, but studies in regard to the constancy of IQ (33, 49, 58, 66) clearly indicate that the processes measured at different age levels are at least closely re-

lated, otherwise it would be impossible to account for the high correlations between the retests.

The Need for the Present Study of the Problem

The need for the present inquiry will be made apparent by an analysis of previous investigations of a similar nature. The results of these studies have already been reviewed in the first chapter.

In Willis' (133) study, no consideration has been given to the question of the variation of IQ with age level. Therefore, the observed median difference of 4.5 points in favor of the second child may very well be a result of differences in the difficulty of the scale.

His findings are further thrown into doubt by Arthur's (5) investigation which shows a discrepancy between the results for the native and those for the immigrant stock, an apparent birth-order difference found only in the latter. It has already been mentioned that the latter result is regarded by the author as having nothing to do with the greater language handicap on the part of the elder children, as the inclusion of 179 pairs, the elder of which had a longer period of language training, does not diminish the difference. However, this conclusion is not necessarily valid in view of the following two possibilities: First, the inequality of the scale may have been an operative factor in the additional material since the elder and younger members of the family were tested at different ages; and, secondly, as the author has actually stated that the families in which tests of three or four children are available are selected largely upon the basis of some school difficulty on the part of the elder children, it is likely that this is also true to a certain extent of the

families in which two children have been examined. This situation may also serve to explain her findings regarding the inferiority of the eldest in families of the former sort.

In fact, Arthur has also touched upon the question of the inequality of the different parts of the scale when considering the IQ's of the 92 pairs of siblings, each sibling having been tested at the end of his first year in the kindergarten. This question, however, has been left out of consideration in dealing with other pairs, the members of which were tested at different age levels. The same factor has been mentioned by Commins (28), but without any provision to compensate for possible resulting errors.

The studies of Sutherland and Thomson (119) and of Jones and Hsiao (78) have shown that there is no difference among different birth orders. In the latter publication, attention has been called to the decline of IQ with age and to the importance of taking this factor into account in dealing with the birth order found in the population under investigation. However, these researches are concerned with birth order rather than birth rank, and therefore no attempt has been made to treat the first-born independently. It is possible that the differences between the first and subsequent children, if any, would be concealed owing to the methods of treatment used.

With regard to the study of Thurstone and Jenkins (121), attention may be called to the fact that, aside from the existence of some differences in favor of the earlier-born (which may be insignificant owing to the small number of cases), some of the differences in favor

of the later-born are far from reliable although based upon large samples. Granting that the possible influence of the negative correlation between CA and IQ has been well taken care of, we still have the following possibilities to consider in interpreting the differences found in favor of the later-born. First, as is pointed out by the authors, an improvement in domestic conditions among families of low social status might involve a progressive reduction of environmental handicaps with the children of higher birth orders. Secondly, common sense seems to indicate that, in the mutual association of siblings, the later-born children can derive a greater relative intellectual gain than the earlier-born. If the first-born suffer from any handicap in intellectual stimulation, this is perhaps likely to become manifest in a sample where, owing to the intellectual inferiority of parents, no provision has been made to compensate for this defect. In fact, of the mothers of a population whose average IQ is around 80, such compensatory provisions can hardly be expected. Thirdly, in their material, the average age of the mother at the birth of the first child may have been below the optimum age for reproduction, and therefore it may have become increasingly favorable with the order of pregnancy. Fourthly, it may be that, in a majority of families studied, the first children that were examined were the cases referred to the clinic for advice concerning some handicap on their part, and, once they were sent, their younger sibs were to follow suit either in furnishing data for comparisons or as a matter of routine. If this is the situation, there will undoubtedly result a successive selection in favor of the later-born.

Fifthly, there may be in the experimental population a portion of children of foreign birth or parentage, among whom the elder members are likely to be more handicapped by the language than the younger ones. If present, such differences can have influenced the total result. Sixthly, the incidence of syphilis, which is said to have a differential effect upon the birth orders, may be a quite important factor in this population of inferior intelligence.

As to Miss Steckel's study, although only a brief reference was made to it by these authors, yet the following probabilities may be mentioned: The standardization of the tests on the experimental population itself has ruled out the possibility of a negative correlation between CA and IQ only so far as this population as a whole is concerned. A further elaboration of this procedure, which is needed, would consist in the exclusion of partial families in the material. The importance of this point may be brought to light as follows: Let us assume the upper age limit of her experimental population to be 13 years. Among the 13-year-olds, in addition to a number of the first-born, there must be some second-born, some third-born and some of the higher birth ranks. The first-born of the families represented by the latter groups among the 13-year-olds were naturally left out. It is reasonable to assume that the number of the families with the first-born left out would decrease with descending age levels. Furthermore, because of greater natal intervals in superior stock, it is not improbable that those partial families should contain a larger proportion of better stock than the average family that was entirely included

in the experimental population. Tests standardized on such a population would naturally yield, for ascending age levels, norms which would increasingly exceed what they would have been had the material been free from the selections of the above sort. Such an effort is unlikely to be offset by the existence, in the lower age groups, of super-normal children among the first-born whose siblings are still in the kindergarten or yet to be born; for each of these age groups contains not only the first- and second-born but also many of the higher birth ranks, and therefore the proportion of the first-born in each group will not be large enough to influence seriously the mean of that group. Now, as to the families that are entirely included in the material, there is no reason to doubt that the earlier-born will fall more frequently in the higher than in the lower age groups. Since it is more difficult for children of the same intelligence to attain the norm at higher than at lower age levels, and since larger numbers of the earlier-born who are to be used for comparison are found in higher age groups, a spurious difference in favor of the later-born would naturally be expected as a result. Therefore, unless the standardization is based only upon the families that are entirely included in the experimental population, the negative correlation between CA and IQ may still arise within each birth order, and this correlation will naturally create a spurious difference in favor of the later-born. As very little has yet been published regarding Miss Steckel's study, we have to suspend our judgment on her results until further information is available. However, apart from the factor of the negative correlation between CA

and IQ, there are other factors to be determined, such as the mother's age at the birth of the first child, the intellectual status of the mother, etc. So long as the material is complicated by any of these factors, it is unsafe to attribute the differences found in favor of the later-born to the influence of birth order, for the other factors are not the necessary concomitants of it.

From the preceding paragraphs, certain general facts are deducible. First, there is lack of agreement among the previous findings regarding the relationship between the order of birth and the intelligence of the child. Secondly, some of the prior investigations are concerned with the comparative status of the elder and younger sibs rather than with that of the first and subsequent children. Differences of the latter sort, if any, could have been concealed owing to the techniques used. Thirdly, the question of the inequality of scale has been either left out of consideration or inadequately attacked. Fourthly, the possibility of some selection against the earlier-born on account of testing or other conditions has failed to be carefully considered. Fifthly, all the differences found in favor of the younger members of the family, assuming them to exist, are possibly attributable to factors other than that of birth order. All these facts seem to indicate the need for a restudy of the problem.

The Nature of the Data

The data used in this study fall into four categories according to the tests used: The Stanford-Binet, the Terman Group Test, the National Intelligence Test, and the Abbreviated Stanford-Binet (starred tests).

computed for each chronological age from 7 to 16 years is .96.

In Table 16 are presented the data classified according to the size of family. This dichotomizing of the material into families of two and of more than two instead of giving the actual size of the latter is based upon the fact that a number of the larger families may still be incomplete. The question of the incompleteness of families, however, does not have to be considered in the present study since it is concerned with the comparison of the first and second children and not with that of the eldest and the youngest members of the family.

In Table 17 are shown the proportions of the male and female sexes among the first- and second-born who are to be compared. It will be noted that between the two birth orders, there is no marked difference in the relative proportions of the sexes.

The chronological ages of the subjects to be considered are those at the time of testing, for it is their IQ's obtained at that time that are to be used for comparison. The term "chronological age" will be abbreviated to CA throughout this study. The CA ranges of the pairs are as follows: (1) 65-169 months, and (2) 55-149 months in the *S.B.* material; (1) 125-244 months, and (2) 115-244 months in the *T.G.* material; and (1) 75-169 months, and (2) 75-154 months in the *N.I.T.* material. The (1) denotes the first sibs and (2) the second. As to the CA range of the singles, *S.B.* cases vary from 60 to 144 months among the first-born and from 35 to 144 months among the second, and *T.G.* cases vary from 110 to 229 among the first and 130 to

249 among the second. The age range of the *N.I.T.* singles cannot be determined owing to the absence of age data in certain cases.

Table 18 gives the CA's of the subjects in terms of central tendencies and variabilities. It should be noted that the average CA of the first-born exceeds that of the second in all kinds of material except that of the *T.G.* singles where approximate equality exists.

The IQ range of the pairs is as follows: (1) 70-159 and (2) 65-154 in the *S.B.* data, (1) 75-149 and (2) 75-154 in the *T.G.* data, (1) 65-139 and (2) 60-159 in the *N.I.T.* data, and (1) 60-149 and (2) 65-149 in the *A.S.B.* data. The IQ range of the singles is a follows: (1) 60-139 and (2) 70-144 in the *S.B.* data, and (1) 75-149 and (2) 70-154 in the *T.G.* data.

Methods of Treatment and Results

Correlations have been computed between the IQ's, on the same tests, of the first and second sibs. In doing so, two methods were used, which may be labelled the method of single entry, and the method of double entry, respectively. According to the method of single entry,

TABLE 16

Data Classified According to the Size of Family

	Birth order							
	1				2			
Kind of material	Two children		More than two		Two children		More than two	
S.B. pairs	92	69%	41	31%	92	69%	41	31%
S.B. singles	42	65%	23	35%	122	67%	71	33%
T.G. pairs	80	44%	103	56%	80	44%	130	56%
T.G. singles	204	67%	102	33%	157	60%	106	40%
N.I.T. pairs	66	56%	52	44%	66	56%	52	44%
N.I.T. singles	91	66%	47	34%	55	53%	49	47%
A.S.B. pairs	82	82%	18	18%	82	82%	18	18%

TABLE 17

Data Classified According to Sex Composition

Kind of material	Birth order 1 Male		Female		2 Male		Female	
S.B. pairs	76	57%	57	43%	71	53%	62	47%
S.B. singles	36	55%	29	45%	99	51%	96	49%
T.G. pairs	93	51%	90	49%	93	51%	90	49%
T.G. singles	153	50%	153	50%	125	48%	138	52%
N.I.T. pairs	71	60%	47	40%	54	46%	64	54%
N.I.T. singles	64	50%	64	50%	53	51%	51	49%
A.S.B. pairs	55	55%	45	45%	54	54%	46	46%
Total	548	53%	485	47%	549	50%	547	50%

TABLE 18

The Chronological Ages of Subjects at the Time of Testing

(Ages in terms of months)

Test	No. cases	Birth order	Mean		Median		*S.D.*	
S.B. pairs	133	1	99.3	2.1	93.4	2.6	24.1	1.5
	133	2	80.9	1.6	74.01	2.0	18.1	1.1
SB. singles	65	1	95.9	2.9	89.7	3.6	23.04	1.1
	193	2	90.2	2.6	84.4	3.2	20.8	1.1
T.G. pairs	183	1	171.5	1.3	170.3	1.6	17.7	.9
	183	2	158.4	1.3	159.6	1.6	17.0	.9
T.G. singles	306	1	163.2	1.04	162.4	1.3	18.2	.7
	263	2	166.5	1.3	163.7	1.6	21.1	.9
N.I.T. pairs	118	1	123.8	1.6	124.3	1.9	16.9	1.1
	118	2	117.4	1.2	116.1	1.6	13.5	.9
A.S.B. pairs	100	1	110.0	1.9	113.9	2.4	18.8	1.3
	100	2	89.5	1.9	88.0	2.3	18.7	1.3

the IQ's of one of the two birth orders are always to be treated as X's and those of the other as Y's. The method of double entry consists in entering each one of a pair of IQ's twice, once as an X and once as a Y. The principle of the latter method is to eliminate the effect of systematic differences, if any, between the first and second sibs so that the result will be a correlation

existing between them without regard to their birth order. However, in determining the reliability of differences between the first and second sibs, the effect of birth order has to be considered, and therefore the correlation obtained by the method of single entry is the one to be used.

The correlations between the sibs are shown in Table 19. The adjusted IQ's refer to the IQ's of the two birth orders that have been adjusted for the same CA. The procedure of adjusting the IQ's will be described later. The symbol "*o*" stands for "original," and the symbol "*c*" stands for "corrected for age." Thus "1*o*" refers to the original IQ of the first-born, and "2*c*" to the corrected IQ of the second-born. Other symbols of the same nature may be similarly inferred. The *r*'s and *P.E.*'s are corrected to the second decimal from the third, which was included in the computation. The double entry method was not used with the *A.S.B.* material for the reason that there was found to be no systematic difference between the two birth orders; and, for the same reason, the adjusted IQ's were not subjected to this kind of treatment.

Correlations have also been computed between IQ and CA for the first- and second-born separately, and the results are given in Table 20. There, one will note that, except in the material of *A.S.B.* pairs, all the negative correlations are reliable. This means that, irrespective of the order of birth, the older the child was at the time of testing, the lower his IQ was likely to be.

In Table 21 are shown the central tendencies and variabilities of IQ's for the first and second sibs respectively, and the reliabilities of their differences in cen-

TABLE 19

CORRELATION BETWEEN THE IQ'S OF THE FIRST AND SECOND SIBS

Test	No. pairs	Original IQ's: r by systematic entry $\pm P.E.$	Original IQ's: r by symmetrical entry $\pm P.E.$	Adjusted IQ's: r by systematic entry $\pm P.E.$	
S.B.	133	.56±.040	.52±.048	1*c* x 2*o*	.56±.040
				1*o* x 2*c*	.55±.040
T.G	183	.52±.054	.51±.026	1*c* x 2*o*	.55±.034*
				1*o* x 2*c*	.54±.035
N.I.T.	118	.39±.053	.27±.041	1*c* x 2*o*	.43±.050
A.S.B.	100	.35±.059		1*o* x 2*c*	.45±.049

*In regard to the material of *T.G.* pairs, one pair was thrown out, because lack of information concerning the examination date of its members made it impossible to adjust their IQ's.

TABLE 20

CORRELATIONS BETWEEN CA AND IQ*

Test	Kind of material	Birth order	No.	r $P.E.$
S.B.	pairs	1	133	—.29±.054
		2	133	—.31±.053
	singles	1	65	—.50±.062
		2	193	—.27±.045
T.G.	pairs	1	183	—.31±.045
		2	183	—.49±.038
	singles	1	306	—.41±.032
		2	263	—.40±.034
N.I.T.	pairs	1	118	—.35±.055
		2	118	—.36±.054
A.S.B.	pairs	1	100	—.20±.074
		2	100	—.23±.064

*No r has been computed for the *N.I.T.* singles owing to the incompleteness of the data concerning dates of examination.

tral tendencies are given in Table 22. The differences are obtained by subtracting the central measure of the second sib from that of the first. Owing to the presence of a relationship between the sibs, the sigma of the difference was computed by the following formula:

$$\sigma_{diff.} = \sqrt{\sigma_1^2 + \sigma_2^2 - 2r_{12}\sigma_1\sigma_2}$$

in which σ_1 denotes the sigma of the central measure of the first sib, σ_2 the sigma of the central measure of the second, and r_{12} the correlation in IQ between the first and second sibs computed by the method of single entry, which expresses the relationship between them taking account of the existence of any systematic differences.

It should be borne in mind that standard errors are used throughout the study except where the symbol "*P.E.*" is given. Here mention may also be made of the fact that the decimals of all the measures (except in the case of r's and critical ratios and in the case where the first decimal is a zero) are corrected to the first place from the second, although the computation has been carried out to three places.

Similar results concerning the singles are presented in Tables 23 and 24. But the sigma of the difference was obtained by the following formula: $\sigma_{diff.} = \sqrt{\sigma_1^2 + \sigma_2^2}$ as no correlation exists between members of different families.

Considering Table 22, one will find that the differences between the first and second sibs in central tendencies are reliable in the case of *S.B.* and *N.I.T.*, but not in the case of *T.G.* and *A.S.B.* As to singles, it is shown in Table 24 that all of the differences found lie within the limits of chance.

However, even the reliable differences are reliable only in appearance in consideration of the following facts: First, as pointed out above, there are negative correlations between CA and IQ within each birth order (see Table 20). This means that, regardless of

TABLE 21

THE CENTRAL TENDENCIES AND VARIABILITIES OF IQ'S AMONG THE FIRST AND SECOND MEMBERS OF THE PAIRS

Test	No. pairs	Birth order	Mean	Median	*S.D.*
S.B.	133	1	102.8 ±1.3	100.2±1.6	15.1±0.9
		2	107.6 ±1.3	108.5±1.5	15.1±0.9
T.G	183	1	107.6 ±1.0	106.9±1.2	13.0±0.7
		2	109.04±1.02	107.9±1.3	13.9±0.7
N.I.T.	118	1	102.1 ±1.3	102.1±1.6	14.2±0.8
		2	106.2 ±1.4	107.5±1.8	15.6±0.9
A.S.B.	100	1	107.5 ±1.3	108.2±1.7	13.3±0.9
		2	109.1 ±1.2	108.6±1.5	12.0±0.8

TABLE 22

DIFFERENCES BETWEEN THE FIRST AND SECOND SIBS AND RELIABILITIES OF THE DIFFERENCES

Test	Measure of central tendency	Diff.	$\sigma_{diff.}$	$\frac{Diff.}{\sigma_{diff.}}$	Chances in 100
S.B.	Mean	—4.9	1.2	3.99	100
	Median	—8.2	1.5	5.57	100
T.G.	Mean	—1.5	1.0	1.49	93
	Median	—1.0	1.2	.82	79
N.I.T.	Mean	—4.1	1.6	2.61	99
	Median	—5.4	1.9	2.85	100
A.S.B.	Mean	—1.5	1.4	1.07	86
	Median	— .4	1.8	.22	58

TABLE 23

THE CENTRAL TENDENCIES AND VARIABILITIES OF IQ'S AMONG THE SINGLES

Test	Birth order	No.	Mean	Median	*S.D.*
S.B.	1	65	102.42±2.21	102.2±3.6	17.9±1.6
	2	193	108.7 ±1.1	108.4±3.2	15.5±0.8
T.G.	1	306	109.5 ±0.8	109.0±1.01	14.2±0.6
	2	263	108.1 ±0.9	107.4±1.1	14.7±0.9
N.I.T.	1	128	105.9 ±1.02	107.0±1.3	17.2±1.1
	2	104	109.8 ±1.4	108.4±1.8	14.7±1.02

corrected $\eta_{IQ.CA}$'s are found to be .35 and .31 for the first- and second-born respectively. Since the $r_{IQ.CA}$'s for the two groups are .29 and .31, the ζ (or difference between η^2 and r^2) obtained for the first-born is .036±.023 and that for the second is 0. These facts show plainly that the assumption of linearity of regression in the *S.B.* material is justifiable. As to the other material, there is not the slightest indication of curvilinearity, and therefore there is no need for testing the linearity of the regression lines.

To illustrate the correction procedure in a concrete manner, let us describe the process of correcting the mean IQ of the first-born in the *S.B.* material. The regression coefficient was obtained by multiplying the negative correlation between CA and IQ found among the first-born (or —.289) by the ratio of the sigma of the distribution of their IQ's to the sigma of the distribution of their CA's (or 15.050/25.050), and it was found to be —.181. Then the difference in average CA between the first- and second-born was obtained by subtracting the average CA of the former from that of the latter (or 80.9—99.3 in months), and it was found to be —18.4 months. By multiplying —.181 by —.18.4, we obtained 3.33, which is the average difference in IQ in favor of the second-born due to the difference in CA in favor of the first (as is computed upon the basis of the regression equation for the first-born). Therefore, in order that the mean IQ's of the two birth orders might be comparable, this average difference in IQ (or 3.33) should be added to the original IQ of the first-born. The result so obtained (or 106.1) is the corrected IQ of the first-born, which is now com-

parable with the original IQ of the second, for both of these IQ's are now based upon the same average age. The process of correcting the mean IQ of the second-born is the same as what is described above except that the average difference in IQ should be subtracted from the original mean IQ of the second-born instead of being added to it.

In this way, the mean of the first-born was corrected on the basis of the average CA of the second, and the mean of the second-born on the basis of the average CA of the first. The standard errors of these corrected means were computed by the following formula (80, p. 117):

$$\sigma_{M_2} = \frac{\sigma_2 \sqrt{1 - r^2}}{\sqrt{N}}$$

It is to be noted that, for the average CA of each birth order, there is an original mean IQ and a corrected mean IQ. The former belongs to the group for which the average CA was originally found, while the latter belongs to the group whose mean IQ has been adjusted to that average CA. In determining the reliability of the difference between the adjusted means, it was necessary to correct the individual IQ's of every pair. The method of doing so was, first of all, to obtain a monthly rate of variation of IQ for each birth order by using the formula:

$$r_{IQ \cdot CA} \quad \frac{\sigma_{IQ}}{\sigma_{CA}}$$

This was used as a unit of increment or decrement according as the CA (at the time of testing) of one mem-

ber of the pair exceeded or was exceeded by that of the other. Take the first-born of an *S.B.* pair for instance. The monthly rate of variation of IQ was found to be .18 points. In the case where the CA of the first-born (at the time of testing) exceeded that of the second by 8 months, 8 units of the monthly IQ variation or 1.44 points were added to the original IQ of the former. But, in the case where the CA of the first-born was exceeded by that of the second by 8 months, the 1.44 points were subtracted from the former's original IQ instead of being added to it. Thus, the IQ of the first sib was adjusted for the CA of the second and the IQ of the second for the CA of the first. Correlations were computed between the IQ's of the pair that had been adjusted for the same age, and therefore there were two correlations to be computed for each group of pairs, one to be used for the difference between the corrected mean IQ of the first sibs and the original mean IQ of the second and the other for the difference between the original mean IQ of the first and the corrected mean IQ of the second.

The results so obtained are presented in Table 25 and Table 26. It is apparent that, except in the case of the comparison of the original mean IQ of the first sibs and the corrected mean IQ of the second in the *T.G.* material, none of the differences are reliable.

We have checked the above findings by taking the difference between the members of each pair in terms of adjusted IQ's and computing the sigma of the average of their differences, and found that the results remain the same.

Now, to attack the problem from another angle, we

TABLE 25

MEAN IQ'S CORRECTED FOR AGE

Test	Kind of material	Av. CA	Birth order	Adjusted mean
S.B.	pairs	80.9	*1c*	106.1±1.2
			2o	107.6±1.3
		99.3	*1o*	102.8±1.3
			2c	102.8±1.2
S.B.	singles	90.2	*1c*	104.6±1.9
			2o	108.7±1.1
		95.9	*1o*	102.4±2.2
			2c	107.5±1.1
T.G.	pairs	158.4	*1c*	110.6±0.9
			2o	109.0±1.02
		171.5	*1o*	107.6±1.0
			2c	103.7±0.9
T.G.	singles	167.5	*1c*	108.1±0.7
			2o	108.1±0.9
		163.2	*1o*	109.5±0.8
			2c	109.3±0.8
N.I.T.	pairs	117.4	*1c*	103.9±1.2
			2o	106.2±1.4
		123.8	*1o*	102.1±1.3
			2c	103.6±1.3

TABLE 26

DIFFERENCES BETWEEN THE ADJUSTED MEANS AND RELIABILITIES OF THE DIFFERENCES

Test	Kind of material	Kind of mean	Diff.	$\sigma_{diff.}$	$\frac{Diff.}{\sigma_{diff.}}$	Chances in 100
S.B.	pairs	*1c* *2o*	—1.5	1.02	1.27	90
		1o *2c*	— .1	1.2	.04	52
S.B.	singles	*1c* *2o*	—4.1	2.2	1.82	96
		1o *2c*	—5.1	2.5	2.07	98
T.G.	pairs	*1c* *2o*	+1.6	.9	1.77	96
		1o *2c*	+3.8	.9	4.22	100
T.G.	singles	*1c* *2o*	+ .1	1.2	.05	52
		1o *2c*	+ .2	1.2	.18	58
N.I.T.	pairs	*1c* *2o*	—2.3	1.4	1.58	94
		1o *2c*	—1.5	1.4	1.06	85

compared the first and second members of each pair, and determined the proportions in which the first sibs are equal, those in which they are superior and those in which they are inferior to the second. The original IQ's, except in the case of *A.S.B.*, were not used because of the existence in the material of the negative correlation between CA and IQ. In comparing the IQ's of each pair, we allowed for all the differences that are less than 3.6 points in the *S.B.* material, all the differences that are less than 2.7 points in the *T.G.* material, all the differences that are less than 4.2 points in the *N.I.T.* material, and all the differences that are less than 3.6 in the *A.S.B.* material. This is based upon the fact that, since the sigmas, in terms of IQ points of the differences between the averages of the two sibs are approximately 1.2, .9, 1.4, and 1.2 in the *S.B.*, *T.G.*, *N.I.T.*, and *A.S.B.* cases, respectively, in order to be reliable, the average difference between the first and second sibs (which is virtually equal to the difference between the averages of the two birth orders) for each kind of material has to be at least three times the sigma of the difference in that material. Therefore, using the average reliable difference as the criterion of comparison, all the differences that fall short of this are to be ignored. It will be noted that, in all these comparisons, the material was classified according to sex and family size, and percentages were computed within each of these classifications on the basis of the total frequencies found under that classification.

The next thing was to determine the difference between the proportion of the cases in which the first sibs are superior and of those in which they are in-

ferior to the second. The differences in favor of the former were labelled +, those in favor of the latter —. In determining such differences, the percentage frequencies of the 1=2 category were equally divided between the 1>2 and 1<2 categories, and then the reliability of the difference between them was computed. Finally, the various kinds of material were combined, and the differences between the 1>2 and 1<2 categories were determined according to the same procedure as is just described.

The findings are shown in Tables 27 to 34. "1=2" means equality between the first- and the second-born, "1>2" means the superiority of the first-born, and "1<2" means their inferiority. It seems that the results of *T.G.* are in favor of the first-born, while those of *S.B.*, *N.I.T.*, and *A.S.B.* are in favor of the second. But in Table 34, one will note that, except the comparisons of the original IQ's of the first sibs with the

TABLE 27

COMPARISON OF ADJUSTED *S.B.* IQ'S BY PAIRS IN TERMS OF ACTUAL FREQUENCIES

Av. CA (months)	Comparative status	Birth order	Two M	Two F	More than two M	More than two F	All sizes M	All sizes F	Total no. pairs
80.9	1=2	1*c*	10	8	4	2	14	10	
		2*o*	9	9	3	3	12	12	24
	1>2	1*c*	17	15	8	6	25	21	
		2*o*	17	15	5	9	22	24	46
	1<2	1*c*	27	15	10	11	37	26	
		2*o*	27	15	10	11	37	26	63
99.3	1=2	1*o*	12	6	4	4	16	10	
		2*c*	9	9	5	3	14	12	26
	1>2	1*o*	18	18	8	6	26	24	
		2*c*	19	17	5	9	24	26	50
	1<2	1*o*	24	14	10	9	34	23	
		2*c*	25	13	8	11	33	24	57

corrected IQ's of the second in the *T.G.* material, none of the percentage differences between the 1>2 and the 1<2 categories are adequately reliable. Moreover, when the total material is considered there is no reliable difference in favor of either category. In all tables, results are classified by sex and by size of family.

TABLE 28

COMPARISON OF ADJUSTED *S.B.* IQ'S BY PAIRS IN TERMS OF PERCENTAGES

Av. CA (months)	Comparative status	Birth order	Two		More than two		All sizes		Total no. pairs
			M	F	M	F	M	F	
80.9	1=2	1*c*	56	44	67	33	58	42	
		2*o*	50	50	50	50	50	50	18
	1>2	1*c*	53	47	57	43	54	46	
		2*o*	53	47	36	64	48	52	35
	1<2	1*c*	64	36	48	52	59	41	
		2*o*	64	36	48	52	59	41	47
99.3	1=2	1*o*	67	33	50	50	62	38	
		2*c*	50	50	63	37	54	46	20
	1>2	1*o*	50	50	57	43	52	48	
		2*c*	53	47	36	64	48	52	38
	1<2	1*o*	63	37	53	47	60	40	
		2*c*	66	34	42	58	58	42	42

TABLE 29

COMPARISON OF ADJUSTED *T.G.* IQ'S BY PAIRS IN TERMS OF ACTUAL FREQUENCIES

Av. CA (months)	Comparative status	Birth order	Two		More than two		All sizes		Total no. pairs
			M	F	M	F	M	F	
166.6	1=2	1*c*	4	5	5	7	9	12	
		2*o*	6	3	5	7	11	10	21
	1>2	1*c*	22	23	25	20	47	43	
		2*o*	20	25	24	21	44	46	90
	1<2	1*c*	12	14	25	20	37	34	
		2*o*	13	13	25	20	38	33	71
173.5	1=2	1*o*	4	8	7	6	11	14	
		2*c*	9	3	8	5	17	8	25
	1>2	1*o*	24	23	27	23	51	46	
		2*c*	22	25	26	24	48	49	97
	1<2	1*o*	10	11	21	18	31	29	
		2*c*	8	13	20	19	28	32	60

TABLE 30

COMPARISON OF ADJUSTED *T.G.* IQ'S BY PAIRS IN TERMS OF PERCENTAGES

Av. CA (months)	Comparative status	Birth order	Two M	Two F	More than two M	More than two F	All sizes M	All sizes F	Total no. pairs
166.6	1=2	1c	44	56	42	58	43	57	12
		2o	67	33	42	58	52	48	
	1>2	1c	49	51	56	44	52	48	
		2o	44	56	53	47	49	51	49
	1<2	1c	46	54	56	44	52	48	
		2o	50	50	56	44	53	47	39
173.5	1=2	1o	33	67	54	46	44	56	
		2c	75	25	62	38	68	32	14
	1>2	1o	51	49	54	46	53	47	
		2c	47	53	52	48	49	51	53
	1<2	1o	48	52	54	46	52	48	
		2c	38	62	51	49	47	53	33

TABLE 31

COMPARISON OF ADJUSTED *N.I.T.* IQ'S BY PAIRS IN TERMS OF ACTUAL FREQUENCIES

Av. CA (months)	Comparative status	Birth order	Two M	Two F	More than two M	More than two F	All sizes M	All sizes F	Total no. pairs
117.4	1=2	1c	6	9	10	1	16	10	26
		2o	6	9	5	6	11	15	
	1>2	1c	8	7	12	10	20	17	
		2o	9	6	8	14	17	20	37
	1<2	1c	24	12	11	8	35	20	
		2o	20	16	6	13	26	29	55
123.8	1=2	1o	6	7	9	1	15	8	
		2c	7	6	5	5	12	11	23
	1>2	1o	9	6	14	10	23	16	
		2c	8	7	8	16	16	23	39
	1<2	1o	23	15	10	8	33	23	
		2c	20	18	6	12	26	30	56

TABLE 32

COMPARISON OF ADJUSTED *N.I.T.* IQ'S BY PAIRS IN TERMS OF PERCENTAGES

Av. CA (months)	Comparative status	Birth order	Two M	Two F	More than two M	More than two F	All sizes M	All sizes F	Total no. pairs
117.4	1=2	*1c*	40	60	91	9	62	38	
		2o	40	60	46	54	42	58	22
	1>2	*1c*	53	47	55	45	54	46	
		2o	60	40	36	64	46	54	31
	1<2	*1c*	67	33	58	42	64	36	
		2o	56	44	32	68	47	53	47
123.8	1=2	*1o*	46	54	90	10	65	35	
		2c	54	46	50	50	52	48	19
	1>2	*1o*	60	40	58	42	59	41	
		2c	53	47	33	67	41	59	33
	1<2	*1o*	61	39	56	44	59	41	
		2c	53	47	33	67	46	54	48

TABLE 33

COMPARISON OF *A.S.B.* ORIGINAL IQ'S BY PAIRS IN TERMS OF ACTUAL FREQUENCIES AND PERCENTAGES

Meaning of figures	Comparative status	Birth order	Two M	Two F	More than two M	More than two F	All sizes M	All sizes F	Total no. pairs
Actual frequencies	1=2	1	7	9	1	2	8	11	
		2	6	10	2	1	8	11	19
	1>2	1	19	6	4	6	23	12	
		2	19	6	5	5	24	11	35
	1<2	1	18	19	6	3	24	22	
		2	19	18	3	6	22	24	46
Percentages	1=2	1	44	56	33	67	42	58	
		2	37	63	67	33	42	58	19
	1>2	1	76	24	40	60	66	34	
		2	76	24	50	50	69	31	35
	1<2	1	49	51	67	33	52	48	
		2	51	49	33	67	48	52	46

TABLE 34

THE RELIABILITY OF DIFFERENCES BETWEEN THE PERCENTAGE FREQUENCIES OF THE 1>2 AND THE 1<2 CATEGORIES

Test	Kind of IQ	Categories compared	%	$\sigma_{\%}$	Diff.	$\sigma_{diff.}$	$\frac{Diff.}{\sigma_{diff.}}$	Chances in 100
S.D.	*1c 2o*	1>2	44.0	4.3	—12.0	6.1	1.97	98
		1<2	56.0					
	1o 2c	1>2	48.0					
		1<2	52.0	4.3	— 4.0	6.1	.65	74
T.G.	*1c 2o*	1>2	55.0					
		1<2	45.0	3.7	+10.0	5.2	1.91	97
	1o 2c	1>2	60.0					
		1<2	40.0	3.6	+20.0	5.1	3.89	100
N.I.T.	*1c 2o*	1>2	42.0					
		1<2	58.0	4.5	—16.0	6.4	2.48	99
	1o 2c	1>2	42.5					
		1<2	57.5	4.6	—15.0	6.4	2.33	99
S.D.+ T.G.+ N.I.T.	*1c 2o*	1>2	48.0					
		1<2	52.0	2.4	— 4.0	3.4	1.17	87
	1o 2c	1>2	51.0					
		1<2	49.0	2.4	+ 2.0	3.4	.58	72
A.S.B.	*1o 2o*	1>2	44.5					
		1<2	55.5	5.0	—11.0	7.1	1.56	94

We have also combined the results obtained in the different kinds of material by getting a weighted average of the differences between the mean IQ's of the first- and second-born that have been adjusted for the same age. Now, for each group of cases, singles as well as pairs (except for the singles of *N.I.T.* and pairs of *A.S.B.*), there are two kinds of differences, the differences between the original mean IQ of the first-born and the corrected mean IQ of the second. It is the differences of the same kind that are to be averaged.

As to the method of averaging the differences, it may be stated as follows: The standard errors of the dif-

ferences are first squared. Then the largest of these squared products is to be divided by itself and by each of the others, and these quotients form a series of weights for the corresponding differences. Naturally, 1 is the weight corresponding to the difference with the largest standard error. The sum of these weighted differences is then to be divided by the sum of the weights, and the quotient so obtained is the weighted average of the differences. This is the method described by Scarborough (110). To give a concrete example, let us suppose that there are two measures to be averaged, which are 100±6 and 80±3. Six squared (or 36) is greater than 3 squared (or 9), and is therefore to be divided by the latter. The resulting quotient (or 4) is the weight for 80, and 1 is the weight for 100. The sum of the weighted measures (or 420) is now divided by the sum of the weights (or 5), and the quotient so obtained (or 84) is the weighted average of the two measures.

In connection with this method, there are two points to be noted. First, where Scarborough mentioned probable errors, standard errors were used instead in this study, but the same results are of course obtainable, as can be easily seen. Secondly, this method differs only in appearance from what Kelley regards as the method of obtaining "best weighted averages" (80, p. 423). According to Scarborough's method:

$$w_1 r_1 = w_2 r_2 = w_3 r_3 = \cdots w_n r_n = C$$

r_1, r_2, etc. being the probable errors of various measures and w_1, w_2, etc. their corresponding weights; the C equals the square of the largest probable error. Therefore, the principle involved here is also to weight each

measure inversely as the square of its probable (or standard) error, and the sum of the weights is also equal to one. The only difference between the methods of Kelley and Scarborough is that the common numerator is 1 in the former case, while it is the square of the largest standard error in the latter.

As to the standard error of this weighted average, it was obtained by the following formula (110):

$$r_0 = \frac{r}{\sqrt{\Sigma w}}$$ [1]

in which r is the largest standard error and Σw is the sum of weights.

The weighted averages of differences are given in Table 35. It will be noted that, in whatever combinations, none of the differences found, plus or minus, are reliable.

Here a word may be said concerning the question of variability. The coefficients of variability are shown in Table 36. *There no marked difference is observable* between the first- and second-born except in the material of *S.B.* and *N.I.T.* singles. The comparability of the two birth orders in the latter two kinds of material is questionable as can be inferred from the above descriptions.

As material for a subsidiary study, we have 427 pairs of first and second sibs whose exact grade position is determinable on our data sheets. The grades studied range from 1L to 8H. Most of these cases are unsuitable for our main study on account of the absence

[1]Thanks are due to Mr. H. S. Conrad for suggesting this formula.

TABLE 35

WEIGHTED AVERAGES OF THE DIFFERENCES BETWEEN THE CORRECTED MEAN IQ'S OF THE FIRST- AND SECOND-BORN

Test and kind of material	Differences found between	Weighted average
S.B. pairs + *S.B.* singles	1*c* and 2*o*	—2.10±1.1
	1*o* and 2*c*	—1.02±1.1
T.G. pairs + *T.G.* singles	1*c* and 2*o*	+1.02±0.7
	1*o* and 2*c*	+2.20±0.7
S.B. pairs + *S.B.* singles + *T.G.* pairs + *T.G.* singles	1*c* and 2*o*	+0.02±0.6
	1*o* and 2*c*	+1.04±0.6
S.B. pairs + *S.B.* singles + *T.G.* pairs + *T.G.* singles + *N.I.T.* pairs	1*c* and 2*o*	—0.30±0.6
	1*o* and 2*c*	+1.00±0.5
S.B. pairs + *S.B.* singles + *T.G.* pairs + *T.G.* singles + *N.I.T.* pairs + *A.S.B.* singles	1*c* and 2*o*	—0.10±0.5
	1*o* and 2*c*	+1.04±0.5

TABLE 36

COEFFICIENTS OF VARIABILITY

Test	Birth order: Pairs 1	Pairs 2	Singles 1	Singles 2
S.B.	14.6	14.0	17.5	14.3
T.G.	12.1	12.7	12.9	13.6
N.I.T.	13.9	14.7	16.2	13.4
A.S.B.	12.4	11.0		

TABLE 37

THE SEX COMPOSITION AND THE FAMILY SIZE OF THE DATA USED FOR COMPARISON ON THE BASIS OF GRADE DEVIATIONS

Birth order	Family size: Two M	Two F	More than two M	More than two F	All sizes M	All sizes F	Total no. pairs
1	119	119	108	81	227	200	427
2	135	103	88	101	223	204	
No. of families	238		189		427		

of test scores. None of the cases in which grades are indicated merely by numbers without specifying as to high or low have been included.

In Table 37 are shown the data classified according to family size and sex composition considering the families of all sizes; there is almost no difference in sex ratio between the two birth orders.

As the basis of comparison in this material consists in the degree of deviation from the normal grade, some system of scoring was needed. For this purpose, the following scale was adopted:

Grade	Years	Grade	Years
1L	6	5L	10
1H	6.5	5H	10.5
2L	7	6L	11
2H	7.5	6H	11.5
3L	8	7L	12
3H	8.5	7H	12.5
4L	9	8L	13
4H	9.5	8H	13.5

On the left-hand side of the scale are the grades, and on the right-hand side of it are the normal ages (in terms of years) of the children who are in the corresponding grades. According to the above scale, if a child of six is in 1H, he is one-half year ahead of his normal grade, and therefore will be given a deviation score of .5. If his sib is eight years old, and is still in 2L, he is one year behind his normal grade, and will receive a deviation score of —1. The difference between the two sibs then is 1.5 points. For the sake of convenience, we may abbreviate the term "grade deviation" to *GD,* which will be used in later paragraphs.

It will be noted that, crude as it is, the scale has the

following point to commend itself. Whatever changes may be made in the upper and lower limits of the age range, the relative position of the sibs will remain constant so long as the ages are separated by half-year intervals. The relative position of the sibs is what we are to consider in this study. As is stated above, the material was selected on the basis of grades rather than on the basis of ages. This was done to facilitate the elimination of cases in which grades were not specified as to high or low and to avoid the difficulty of scoring the children who were recorded as being in the kindergarten.

The ages of subjects were computed up to the beginning of the semester in which their grades were reported. Any period of life amounting to less than a half-year was neglected. This was to allow for the fact that grade deviations of less than .5 point might have been merely accidental.

The results are presented in Tables 38 to 40. Table 38 seems to indicate the existence of some difference between the first and second sibs; but, considering the method of computing ages that was used, differences of less than .5 point are not to be regarded as significant.

Moreover, in this material, there exists a negative correlation between *CA* and *GD*. The correlation between these two variables has been computed for the two birth orders separately. It is found to be —.44± .026 among the first sibs and —.29±.030 among the second. This means that, regardless of whether the child is the first or the second member of the family, the older he is, the more likely he is to stay behind his normal grade.

Now, since the average CA of the first sibs exceeds that of the second, we had to resort to the same procedure of adjustment as is described in connection with the study of IQ's. When the mean *GD* scores have been so adjusted for age, the differences, as are shown in Table 41, turn in favor of the first sibs, but these differences amount to less than .5 point, and are therefore negligible.

However, in the respect of variability, the first sibs surpass the second. This, as we have seen, is true only of the grade deviations.

TABLE 38

COMPARISON OF THE FIRST AND SECOND SIBS ON THE BASIS OF GRADE DEVIATIONS

Birth order	Mean	Median	*S.D.*
1	—.1±.05	+.02±.10	1.0±.03
2	+.1±.04	+.20±.05	.8±.03

TABLE 39

THE CHRONOLOGICAL AGES OF THE TWO BIRTH ORDERS

(In terms of years)

Birth order	Mean	Median	*S.D.*
1	12.5±.1	12.4±.1	1.8±.1
2	10.1±.1	10.1±.1	2.1±.1

TABLE 40

MEAN GRADE DEVIATIONS ADJUSTED FOR AGE

Av. CA (years)	Birth order	Mean
10.067	1*c*	+.5±.04
	2*o*	+.1±.04
12.470	1*o*	—.1±.05
	2*c*	—.1±.04

It is to be understood that the study of grade deviations is included here merely as a subsidiary approach to the problem and that the results derived therefrom can be expected to throw light upon it only insofar as the grade position is a reliable index of intelligence.

Summary and Conclusion

Previous studies of the relationship between the order of birth and the intelligence of the child have not yet furnished any decisive evidence concerning the handicapping of the first-born. The findings so far available are not only in mutual disagreement, but the differences found in favor of the later-born are attributable to causes other than that of birth order. Hence the need for a further study of the problem.

In the present inquiry, the comparison is limited to the first two birth ranks. This is justified by the fact that, if the primogeniture should tend to be associated with intellectual handicap, the first-born would be found to be reliably inferior to the second as well as to any other subsequent children. The data consist of the test results of the Stanford Revision of the Binet Scale, the Terman Group Test of Mental Ability, the National Intelligence Test and the Abbreviated Stanford-Binet. A total number of 2127 cases have been included.

In order to take care of the possible divergencies in the hereditary factors of different families, the procedure is to compare not only the first- and second-born in general, but also the first and second members of the same family. In fact, the latter comparison is where our emphasis lies. In the first three kinds of material

mentioned above, it is found that the mean IQ difference, wherever it exists, is in favor of the first-born or the second-born according as the average CA of the latter exceeds or is exceeded by that of the former, but the elimination of their average discrepancy in CA results in the removal of their mean difference in IQ. In the case of the *A.S.B.* material, the difference, even in terms of original IQ's, is found to be unreliable. Therefore, we may conclude that there is no difference in intelligence between the first- and second-born.

This conclusion is supported by a study of the grade deviations of the first and second sibs insofar as the grade position is an index of intelligence.

This failure to find the alleged difference in intelligence between the first- and second-born does not seem to be due to any of the following causes:

1. It is not due to factors associated with family size (16, 26, 32, 40, 79, 83, 119), for these have been ruled out in the comparison of siblings.

2. It is not due to the possibility that the results have been affected by a difference in sex ratio between the first- and second-born; for, except in the material of *N.I.T.* pairs, there is no marked difference in sex ratio, and in Tables 27 to 33, sex does not appear to play a part in the determination of the three categories, $1=2$, $1>2$, and $1<2$.

To make a more thorough analysis of the possible bearings of sex differences upon the problem, each kind of test material was classified into four types of sibships: (*a*) both males, (*b*) both females, (*c*) first male and second female, and (*d*) first female and second male, and the average of the birth-order differences

for each type of sibship was computed. Here the original, instead of the adjusted IQ's, were used. This was done on the ground that, in case of the existence of a systematic difference between the sexes, one of the sexes would have been overcorrected and the other undercorrected. The results obtained are presented in Table 42. It will be noted that there is no systematic difference in the respect of birth-order differences among the various types of siblings.

TABLE 41

AVERAGES OF BIRTH-ORDER DIFFERENCES CLASSIFIED ACCORDING TO SEX

(In terms of original IQ's)

	Kind of material			
sibship	*S.B.*	*T.G.*	*N.I.T.*	*A.S.B.*
M_1 *vs.* M_2	−6.5±.4	−3.9±.4	−5.4±.6	+4.6±.6
F_1 *vs.* F_2	−2.3±.6	−4.7±.5	−2.9±.7	−2.0±.6
M_1 *vs.* F_2	−5.1±.5	−2.3±.3	−3.8±.4	−5.0±.4
F_1 *vs.* M_2	−6.0±.5	−1.1±.4	−5.1±.6	−3.0±.7

3. It is not due to errors in the classification of the first-born, arising from the fact that a certain proportion are actually second-born through the occurrence of previous unrecorded abortions, miscarriages, stillbirths and deaths from premature births. From a study of 311 unselected families of two or more children in the same locality from which our material was drawn, the percentages of these prenatal casualties are found to be 15.7 and 15.5 for the first- and second-born, respectively. As the testing program was unselected in nature as far as our *S.B.* and *T.G.* cases are concerned, and as our study deals exclusively with families of two or more children, it seems reasonable to assume that the

same percentages of prenatal casualties obtain in these two kinds of material. In other words, 15.7% of the so-called first-born are second-born, and 15.5% of the so-called second-born are third-born. Now, if intelligence were to increase with the order of birth as Thurstone and Jenkins have claimed, then almost the same proportion would be raised in intelligence among the first-born as among the second-born, and consequently the relative position of the two birth orders would remain the same as if no prenatal casualties at all had occurred. Furthermore, even if we assume that the second-born would exceed the first in intelligence, but would not be exceeded by the third, or if we assume that a certain proportion of the so-called first-born were actually the second-born but none of the so-called second-born were third-born, the proportion of the actual first-born, being more than 5 times as large as that of the pseudo-firstlings, would pull down the mean IQ of their group, so that the difference would still be in favor of the second-born. But our study shows that, when the IQ's have been corrected for age, the difference between the first- and second-born are not only unreliable, but they are not even in the same direction in the various tests. Therefore, we may set aside the possibility of our results being marred by the occurrence of prenatal casualties in the two birth orders.

4. It is not due to the possibility that a significantly larger proportion of the first sibs than of the second were born within the life period supposed to be favorable for reproduction, thus offsetting the alleged advantage of birth order. A study of 254 unselected fam-

ilies of two or more children in the same locality from which our samples were drawn shows that the age of the mother at the first birth ranges from 15 to 35 years with the average at a little over 23. Approximately the same average age may be reasonably assumed for our material. Since the average natal interval between the first and second children of the same family ranges from 25.1 to 27.6 months in the four kinds of data, the ages of the mothers at the second birth are likely to fall around 25. These facts seem to show that there is no ground for believing that the alleged tendency for intelligence to increase with the order of birth was offset by the apparent tendency for the quality of offspring to deteriorate with mother's age after 30 as is shown in certain studies (9, 11, 41, 65, 105, 117, 124, 125).

Moreover, in the *A.S.B.* material, we have secured information concerning the ages of parents at the birth of the children studied in 80% of the cases. When the IQ of the child is correlated with the age of the father and the age of the mother, the coefficients are found to be $-.024 \pm .075$ and $+.10 \pm .074$ in the former case and $+.086 \pm .075$ and $+.23 \pm .071$ in the latter for the first- and second-born respectively. These findings seem to show that, if there is any relationship at all between the IQ of the child and the age of the mother, it is a positive rather than a negative one. Therefore, the hypothesis is untenable that, in this material, the advantages of the second birth have been offset by the influence of the mother's age. It may also be mentioned that, in this sample the fathers' ages range from 20 to 45 for the first-born with an average of 29.0 and an *S.D.* of 5.5, and from 22 to 49 for the second with an aver-

age at 31.2 and an *S.D.* of 5.7; and the mothers' ages range from 17 to 38 for the first-born with an average at 26.1 and an *S.D.* of 4.5, and from 20 to 39 for the second with an average at 28.0 and an *S.D.* of 4.8. The average ages of the mothers agree quite closely with what was inferred from the above-mentioned study of 254 unselected families.

The above facts seem to warrant the conclusion that the first-born are at least not inferior to the subsequent children. But, if the later-born do derive a greater intellectual gain from the association with siblings than the first-born, if the experience of the mother in bringing up the children, and the social and economic status of the family do improve with successive birth orders, and if, in spite of these conditions, the first-born still stand on an equal footing with the later-born, it will be reasonable to assume either that these factors do not exercise an appreciable effect upon their IQ's or that the first-born are actually superior to the children of subsequent ranks but unfavorably influenced by these factors. At any rate the present study tends to weaken the claim that the reduction of family size has a detrimental effect upon the average intelligence of the population; by the time of reaching school (if not before) the first-born are equal to the second-born. It would indeed be surprising if a later study, employing similar methods in a similar population, were to demonstrate that the first two birth orders are inferior to the third and subsequent orders.

REFERENCES

1. Abt, I. A., Adler, H. M., & Bartelme, P. The relationship between the onset of speech and intelligence. *J. Amer. Med. Asso.*, 1929, 93, 1351-1356.
2. Adair, F. L., & Stewart, C. A. Milk ingestion in relation to changes in body weight of newborn infants. *J. Amer. Med. Asso.*, 1922, 78, 1865-1868.
3. Adler, A. Characteristics of the first, second, and third child. *Children*, 1928, 3, 14 and 52.
4. Ansell, C. Statistics of families in the upper and professional classics. London: National Life Assurance Soc., C. & E. Layton, 1874. Pp. 89.
5. Arthur, G. The relation of I.Q. to position in family. *J. Educ. Psychol.*, 1926, 17, 541-550.
6. Auerbach, E. Kurzsichtigkeit und Erstgeburt. *Arch. f. Rassen.- u. Gesellschbiol.*, 1912, 9, 762-763.
7. Baker, H. J., Decker, F. J., & Hill, A. S. A study of juvenile theft. *J. Educ. Res.*, 1929, 20, 81-87.
8. Bateman, W. G. The language status of three children at the same ages. *Ped. Sem.*, 1916, 23, 211-240.
9. Becker, W. H. Was wird aus dem Kindern alter Erstgebärender? *Arch. f. Rassen.- u. Gesellschbiol.*, 1921, 13, 277-297.
10. Beeton, M., & Pearson, K. On the inheritance of the duration of life and on the intensity of natural selection. *Biometrika*, 1901, 1, 50-99.
11. Bell, A. G. The duration of life and the conditions associated with longevity. Washington, D. C.: Geneal, Rec. Off., 1918. Pp. 57.
12. Blanchard, M. B. The applicability of the Stanford Revision of the Binet-Simon Scale of Intelligence to a typically rural group of children. (Unpublished M.A. thesis.) Columbia Univ., 1927.
13. Bloch, O. Langage d'action dans les première stades du langage de l'enfant. *J. de psychol.*, 1923, 20, 670-674.
14. Boas, F. The growth of first-born children. *Science*, 1895, 1, 402-404.
15. Bolt, R. A. The mortalities of infancy. In Abt's *Pediatrics*, Vol. 2. Philadelphia: Saunders, 1926.

16. BRADFORD, E. J. Can present scholastic standards be maintained? *Forum Educ.*, 1925, 3, 186.

17. BRECKENRIDGE, S. P., & ABBOTT, E. The delinquent child and the home. New York: Charities Publ. Comm., 1912. Pp. 355.

18. BREHMER, H. Die Aetiologie der chronischen Lungenschwindsucht. Wiesbaden, 1885.

19. BROUSSEAU, K., & BRAINERD, H. G. Mongolism. Baltimore, Md.: Williams & Wilkins, 1928. Pp. viii+210.

20. BURDGE, H. G. Our boys. Albany: Lyon, 1921. Pp. 59.

21. BUSEMANN, A. Geschwisterschaft, Schultuchtigkeit und Charakter. *Zsch. f. Kinderforsch.*, 1928, 34, 1-52.

22. CADRE, M. Étude sur les modifications du poids, de la taille et de la suture sagitalle chez le nouveau-né (dans les 9 premiers jours). Paris, 1910. Pp. 120.

22*a*. CAMPBELL, H. Causation of disease.

23. CARMAN, A. Pain and strength measurement of 1507 school children in Saginaw, Michigan. *Amer. J. Psychol.*, 1899, 10, 392-398.

24. CASTELL, A. Über den Einfluss des Alters der Mutter auf Gewicht und Lange des neugeborenen Kindes. Königsberg, 1869. Pp. 31.

25. CATTELL, J. M. American men of science. (4th ed.) New York: Science Press, 1927. Pp. viii+1132.

26. CHAPMAN, J. C., & WIGGINS, D. M. Relation of family size to intelligence of offspring and socio-economic status of family. *Ped. Sem.*, 1925, 32, 414-421.

27. CHASE, J. H. Weakness of eldest sons. *J. Hered.*, 1914, 5, 209-211.

28. COMMINS, W. D. The intelligence of the later born. *School & Soc.*, 1927, 25, 488-489.

29. COBB, J. A. The alleged inferiority of the first-born. *Eugen. Rev.*, 1913, 5, 358.

30. CRZELLITZER, A. Wie vererben sich Augenleiden. *Ges. f. Soz. Med. Hygiene u. Medizinalstatistik*, 1910.

31. DAYTON, N. A. Order of birth of mental defectives. *J. Hered.*, 1929, 20, 219-224.

32. ———. Intelligence and size of family. *J. Hered.*, 1929, 20, 365-374.

33. DICKSON, V. Mental tests and the classroom teacher. Yonkers, N. Y.: World Book Co., 1923. Pp. 231.

34. Down, L. On some of the mental affections in childhood and youth. London, 1887.

35. Dublin, L. J., & Langman, H. On the handicapping of the first-born. A criticism of Prof. Pearson's 1914 memoir. *Publ. Amer. Stat. Asso.*, 1915, 14, 727-735.

36. Dugdale, R. L. The Jukes. (4th ed.) New York: Putnam, 1910. Pp. 120.

37. Duncan, J. M. On the weight and length of the newly-born child in relation to mother's age. *Edinburgh Med. J.*, 1864, 10, 497-502.

38. ———. The laws of the fertility of women. *Trans. Roy. Soc.*, 1866, 24, 287-314.

39. Duncan, M. *Lancet*, 1883 (Jan. and March).

40. Ellis, H. A study of British genius. (New. ed., rev. and enlarged.) Boston: Houghton Mifflin, 1926. Pp. xvi+396.

41. Ewart, R. J. The influence of parental age on offspring. *Eug. Rev.*, 1911, 3, 201-232.

42. ———. The influence of age of the grandparent on the number of children born and their sex. *J. Hygiene*, 1915, 15, 160.

43. Fasbender, H. Mutter and Kindeskörper. Das Becken des lebenden Neugeborenen. *Zsch. f. Geburtschülfe u. Gynäkol.*, 1878, 3, 278-304.

44. Fourmann, F. Wovon ist das Gewicht der Neugeborenen abhängig? Bonn: Carl Georgi, 1901. Pp. 68.

45. Freeman, F. N., *et al.* The influence of environment on the intelligence, school achievement and conduct of foster children. *27th Yrbk. Nat. Soc. Stud. Educ.*, 1928, Part 1, 102-217.

46. Gale, M. C., & Gale, H. A. The vocabularies of three children in one family at two and a half years of age. *Psychol. Stud.*, 1900, 1, 70-117.

46a. ———. The vocabularies of three children in one family at two and three years of age. *Ped. Sem.*, 1902, 9, 422-435.

47. Galton, F. English men of science: their nature and nurture. London: Macmillan, 1874. Pp. xiii+270.

48. ———. Hereditary genius. New York: Appleton, 1887. Pp. xii+390.

49. Garrison, S. C. Additional retests by means of the Stanford-Binet. *J. Educ. Psychol.*, 1922, 13, 307-312.

GEISSLER, A. Ueber den Einfluss der Sauglingssterblichkeit auf die eheliche Fruchtbarkeit u.s.w. *Zsch. d. Sachs. Stat. Bur.*, 1885.

GHEORGOV, J. A. Ein Beitrag zur grammatischen Entwicklung der Kindersprache. Leipzig: W. Engelmann, 1908. Pp. 191.

————. La vocabulaire dans le langage des enfants. Sofia, 1910. Pp. 301. (L'Annuaire de L'Universite de Sofia. En Bulgare.)

GINI, C. *Rev. ital. di sociol.*, 1914.

————. Superiority of the eldest. *J. Hered.*, 1915, 6, 37-39.

GOODENOUGH, F. L. The emotional behavior of young children during mental tests. *J. Juv. Res.*, 1929, 13, 204-219.

GOODENOUGH, F. L., & LEAHY, A. M. The effect of certain family relationships upon the development of personality. *Ped. Sem.*, 1927, 34, 45-71.

GORING, C. B. The English convict; a statistical study. London: Wyman & Sons, 1913. Pp. 280.

GRAY, P. L., & MARSDEN, R. E. The constancy of the intelligence quotient: further results. *Brit. J. Psychol.*, 1924, 15, 169-173.

GREENWOOD, M., & YULE, G. U. On the determination of size of family and of the distribution of characters in order of birth from samples taken through members of the sibships. *Roy. Stat. Soc. J.*, 1913-1914, 77, 179-197.

GRIFFITH, J., CROZER, P., & GIDDINGS, J. C. The weight of breast-fed infants during the first two weeks of life. *Arch. Ped.*, 1907, 24, 321-345.

HANSEN, S. Über die Minderwertigkeit der erstgeborenen Kinder. *Arch. f. Rassen.- u. Gesellschbiol.*, 1913, 10, 701-722.

HARRIS, J. A. Relationship between pregnancy order and birth order and length and weight of newborn infants. *Proc. Soc. Exper. Biol. & Med.*, 1926, 23, 806-808.

HECKER, C. Über Gewicht und Länge der neugeborenen Kinder in Verhaltnis zum Alter der Mutter. *Monatsschr. f. Geburtskunde u. Frauenkrankheiten*, 1865, 26, 348-363.

HERON, D. A first study of the statistics of insanity and the inheritance of the insane diathesis. (Univ. London

Publs. from the Francis Galton Lab.) London: Dulau, 1907. Pp. 33.

65. Hibbs, H. H., *et al.* Infant mortality among infants classified according to age of mothers. (Investigation of Boston, Mass.) New York: Russell Sage Foundation, 1916.

66. Hildreth, G. Stanford-Binet retests of 441 school children. *Ped. Sem.*, 1926, 33, 265-386.

67. Hillebrand, F. Untersuchungen über die Milch-zufuhr und über die Jodkaliumausscheidung des Säuglings. 1884-1885.

68. Hirsch, N. D. M. An experimental study of the East Kentucky mountaineers. *Genet. Psychol. Monog.*, 1928, 3, 183-244.

69. Holden, F. S. On the vocabularies of children under two years. *Trans. Amer. Philol. Asso.*, 1877, 8, 58-68.

70. Holmes, S. J. The trend of the race. New York: Harcourt, Brace, 1921. Pp. 396.

71. ————. Age at parenthood, order of birth and parental longevity in relation to the longevity of offspring. *Univ. Calif. Publs. Zool.*, 1928, 31, 359-375.

72. Holmes, S. J., & Wilson, I. Age of parents and order of birth in relation to longevity of offspring. *J. Hered.*, 1925, 16, 47-50.

73. Hutchinson, R. Congenital pyloric stenosis. *Brit. Med. J.*, 1910, 2, 1021.

74. Hug-Hellmuth, H. Vom "mittleren Kinde." *Imago*, 1921, 7, 93-94.

75. Issmer, E. Zwei Haupt merkmale der Reife neugeborenen und deren physiologische Schwankungen. *Arch. f. Gynäkol.*, 1887, 30, 277-315.

76. Jansen, M. Weakness in power to grow and congenital dwarfism. *Rev. fran. de ped.*, 1925, 1, 529.

77. Jones, C. E. A genealogical study of population. *Publs. Amer. Stat. Asso.*, 1918, 16, 201-219.

78. Jones, H. E., & Hsiao, H. H. A preliminary study of intelligence as a function of birth order. *J. Genet. Psychol.*, 1928, 35, 428-433.

79. Keller, H. Über die Bezeihungen zwischen Begabung und Fortpflanzung. *Arch. f. Rassen.- u. Gesellschbiol.*, 1929, 22, 36-49.

80. KELLEY, T. L. Statistical method. New York: Macmillan, 1923. Pp. xi+390.

81. KEY, W. E. *J. Hered.*, 1926, 17, 150.

82. KÉZMÁRSZKY, T. Über die Gewichtsveränderungen reifer Neugeborener. *Arch. f. Gynäkol.*, 1873, 5, 547-561.

83. LENTZ, T. Relation of I.Q. to size of family. *J. Educ. Psychol.*, 1927, 18, 486-496.

84. MACAULEY, T. B. Supposed inferiority of first- and second-born members of families: statistical fallacies. *Amer. Breeder's Mag.*, 1911, 2, 165-175.

85. MARCH, L. Some researches concerning the factors of mortality. *J. Roy. Stat. Soc.*, 1912, 75, 513-520.

86. MARRO, A. Influence of age of parents on offspring. *Prob. in Eugen.*, 1912, 118-136.

87. McDONALD, B. *Boston Med. & Surg. J.*, 1901.

88. MITCHELL, A. Some statistics of idiocy. *Edinburgh Med. J.*, 1866, 11, Pt. 2, 639-645.

89. MOLL, A. Untersuchungen über die Libido sexualis. Berlin, 1898. Vol. 1.

90. MOURLOT, C. Des variations de poids chez les nouveaunis nourris par leurs mères, pendant les dix premiers jours. Paris, 1892. Pp. 76.

91. NICE, M. M. Ambidexterity and delayed speech development. *Ped. Sem.*, 1918, 25, 141-162.

92. ———. A child who would not talk. *Ped. Sem.*, 1925, 32, 105-142.

93. ORDAHL, G. Birth rank of Mongolians. *J. Hered.*, 1927, 18, 429-431.

94. PEARSON, K. On telegony in man. *J. Stat. Soc.*, 1897, 60, 440.

95. ———. A first study of the statistics of pulmonary tuberculosis. *Stud. Nat. Deter.*, 1910, 2.

96. ———. The problem of practical eugenics. *Univ. London Eug. Lab. Lecture Series*, 1912, 5.

97. ———. On the handicapping of the first-born. *Univ. London Eug. Lab. Lecture Series*, 1914, 10.

98. PELLER, S. Der Einfluss sozialer Moment auf den Körperlichen Entwicklungszustand der Neugeborenen. *Wien Arb. a. d. Geb. d. soz. Med.* (Vienna and Leipzig), 1913, Sec. 5, 1-47.

99. PLOETZ, A. Zusammenhang der Sterblichkeit der Kinder mit dem Lebensalter der Eltern bei ihrer Geburt und mit der Geburtsreihenfolge. *Arch. f. Rassen.- u. Gesellsch-biol.*, 1911, 8, 761-763.

100. ———. Ziele und Aufgabe der Rassenhygiene. *Dtsch. Vjsch. f. öffentl. Gesundheitspflege*, 1911, 43, 165.

101. ———. Neomalthusianismus und Rassenhygiene. *Arch. f. Rassen.- u. Gesellschbiol.*, 1913, 10, 166-172.

102. POPENOE, P. The long-lived first-born. *J. Hered.*, 1916, 7, 395-398.

103. REDFIELD, C. L. Control of heredity. Chicago: Monarch, 1903.

104. REITER, H. Wird die Sterblichkeit vor vollendeter Aufzucht durch Geschwisterzahl und soziale Lage der Eltern beeinflusst? *Zsch. f. Hygiene u. Infektions-Krankh.*, 1923, 99, 76-83.

105. REITER, H., & IHLEFELDT, H. Einfluss von Alter der Mutter und Geburtigkeit des Kindes auf dessen Entwicklung. *Klin. Woch.*, 1922, 1, 2524-2525.

106. RÉVÉSZ, B. Der Einfluss des Alters der Mutter auf die Körperhöhe. *Arch. f. Anthrop.*, 1906, 32, 160-167.

107. REYNOLDS, B. C. Environmental handicaps of 400 habit clinic children. *Hosp. Soc. Serv.*, 1925, 12, 329-336.

108. RIVERS, W. C. Consumption and order of birth. *Lancet*, 1911, 181, 994-1004.

109. ———. Primogeniture and abnormality: a possible fallacy. *Eug. Rev.*, 1914, 6, 58-61.

110. SCARBOROUGH, J. B. The invalidity of a commonly used method for computing a certain probable error. *Proc. Nat. Acad. Sci.*, 1929, 15, 665.

111. SCHLESINGER, E. Die Kinder der kinderreichen Familien. *Arch. f. Kinderhk.*, 1923, 73, 50-68.

112. SCHMID, M. Über den Einfluss des Militärdienstes der Väter die körperliche Entwicklung ihrer Nachkommenschaft. *Jahrb. f. Kinderhk.*, 1892, 33, 327-350.

113. SCHNEIDER, O. Die Schöpferischekraft des Kindes in der Gestaltung seiner Bewusstseinzustände bis zur Beginn des Schulunterrichts. *Zsch. f. Phil. u. phil. Kritik*, 1903, 121, 153-174.

114. SHUTTLEWORTH, G. E. *Brit. Med. J.*, 1900, 17, 1446.

SMITH, M. E. An investigation of the sentence and the extent of vocabulary in young children. *Univ. Ia. Stud.: Stud. Child Welfare,* 1926, 3, No. 5. Pp. 92.

STERN, C., & STERN, W. Die Kindersprache: eine psychologische und sprachtheoretische Untersuchung. Leipzig: Barth, 1922. Pp. 434.

STRAHAN, S. A. K. Marriage and disease. New York: 1892. Pp. 250.

STRATTON, G. M. Anger and fear: their probable relation to each other, to intellectual work, and to primogeniture. *Amer. J. Psychol.,* 1927, 36, 125-140.

SUTHERLAND, H. E. G., & THOMSON, G. H. The correlation between intelligence and size of family. *Brit. J. Psychol.,* 1926, 17, 81-92.

TERMAN, L. M., *et al.* Genetic studies of genius. Vol. 1. The mental and physical traits of a thousand gifted children. Stanford Univ., Calif.: Stanford Univ. Press, 1925. Pp. xv+648.

THURSTONE, L. L., & JENKINS, R. L. Birth order and intelligence. *J. Educ. Psychol.,* 1929, 20, 641-651.

THURSTONE, L. L., & THURSTONE, T. G. A neurotic inventory. *J. Soc. Psychol.,* 1930, 1, 3-30.

TREDGOLD, A. F. Mental deficiency. London: Baillière, Tindall & Cox, 1922. Pp. xx+569.

VAERTING, M. Das günstigste Zeugungsalter für die geistige Fahigkeit der nachkommen. Würzburg: C. Kabitsch, 1913. Pp. 63. *Neue Generation,* 1914, 10, 94.

———. Das Altersverhältnis der Eheleute. *Neue Generation,* 1914, 10, 497.

VON DER VELDEN, F. Die Minderwertigkeit der Erstgeborenen. *Arch. f. Rassen.- u. Gesellschbiol.,* 1908, 5, 526-530.

WEEKS, D. F. The inheritance of epilepsy. *Prob. Eug.,* 1912. London: Knight, 1912. Pp. 40.

WEINBERG, W. Die rassenhygienische Bedeutung der Fruchtbarkeit. *Arch. f. Rassen.- u. Gesellschbiol.,* 1910, 7, 684.

———. Die Kinder der Tuberkulösen. Leipzig: Hirzel, 1913. Pp. vi+160.

———. Kurzsichtigkeit und Erstgeburt. *Arch. f. Rassen.- u. Gesellschbiol.,* 1913, 10, 326-327.

131. WELLS, G. S. A study of the order of the birth of children. (Unpublished thesis.) Stanford Univ., 1901.

132. WESTERGAARD, H. Die Lehre der Mortalität und Morbidität. Jena: G. Fischer, 1882. Pp. vi+504.

133. WILLIS, C. B. The effects of primogeniture on intellectual capacity. *J. Abn. & Soc. Psychol.*, 1924, 18, 375-377.

134. WINTER, L. *State Hosp. Bull.*, 1897, 463.

135. YODER, A. H. A study of the boyhood of great men. *Ped. Sem.*, 1894, 3, 134-156.

L'ÉTAT DU PREMIER-NÉ, EN RAPPORT AVEC L'INTELLIGENCE

(Résumé)

Les études antérieures ont indiqué l'influence probable de l'ordre de naissance sur des certains traits physiques, comme le montre le compte-rendu critique qui accompagne cet article. Les données sont moins conclusives à l'égard de l'intelligence et des autres traits du comportement. Dans ce rapport-ci, on s'est servi des résultats des tests d'intelligence chez 2127 "siblings" (enfants non uniques) dans une étude comparative de la capacité mentale du premier-né et du deuxième-né. Les sujets ont été élèves des classes (années scolaires) des écoles publiques de Berkeley et d'Oakland, Californie, et l'on a employé comme tests, aux niveaux d'âge qu'il fallait, la Revision Stanford de l'échelle Binet-Simon, le Test Collectif Terman de Capacité Mentale, et le Test National d'Intelligence. Parmi 1068 individus de paires de "siblings," le cadet s'est montré avoir supériorité moyenne du QI de 1,05 à 4,09 points dans les divers tests. Ce résultat s'accorde avec les études récentes de Thurstone et Jenkins, Steckel, Arthur, et d'autres. On doit, cependant, considérer deux faits avant de faire une interprétation de ces résultats: (1) les deuxième-nés, en moyenne, sont de 6 à 18 mois plus jeunes que les premier-nés, à l'époque du test; (2) les inégalités d'échelle, dans les divers tests, impliquent une corrélation négative entre QI et CA, comme type, de l'ordre de —,03 dans chaque ordre de naissance considéré seul. Une correction pour le facteur âge, par l'application de la formule de régression, amène à la conclusion que l'intelligence n'est pas une fonction de l'ordre de naissance, au moins en tant qu'il s'agit des deux premiers ordres de naissance. Ceci est soutenu par une étude du premier-né et du deuxième-né des différentes familles, dérivée des mêmes élèves; on a trouvé à peu près les mêmes QI pour les deux ordres de naissance, après avoir fait les corrections nécessaires. Ceci est soutenu aussi par une étude de l'état scolaire (en termes de position en classe) de 427 paires de "siblings" premier-né et deuxième-né; on n'a trouvé aucunes différences dans la vitesse de leur avancement à l'école.

Ces résultats se sont montrés indépendants de certains variables qui pourraient les déranger, tels que l'âge de la mère, l'effet des accidents avant la naissance, les proportions différentielles de sexe dans les deux ordres de naissance, et les facteurs de sélection causés par les variables associés à la grandeur de la famille.

HSIAO

DIE FÄHIGKEITEN DER ERSTGEBORENEN, MIT BESONDERER BERÜCKSICHTIGUNG DER INTELLIGENZ

(Referat)

Frühere Studien haben gezeigt, dass die Geburtsordnung wahrscheinlich gewisse physische Merkmale beeinflusst. Die beiliegende kritische Beurteilung gibt darüber Rechenschaft. Die Angaben sind weniger überzeugend bezüglich der Intelligenz und anderer Merkmale des Verhaltens. In der hier berichteten Untersuchung wurden die Ergebnisse der Intelligenzteste an 2127 Nachkommen geprüft, und mit Bezug auf die Geistesfähigkeiten der Erst- und Zweitgeboren miteinander verglichen. Das Material wurde in den 12 Stufen der Volksschulen von Berkeley und Oakland, Californien gewonnen, und die den Altersstufen angepassten Teste waren die Stanford

Umarbeitung der Binet-Simon Skala, die Terman- Gruppenteste für Geistesfähigkeiten und der Nationale Intelligenztest. Unter 1068 Individuen von Nachkommenpaaren zeigen die Zweitgeborenen eine durchschnittliche IQ Überlegenheit von 1,5 bis 4,9 nach den verschiedenen Testen. Diese Ergebnisse stimmen mit den den neuern Untersuchungen von Thurstone und Jenkins, Steckel, Arthur und anderen überein. Ehe man aber eine Interpretation der Ergebnisse geben kann, müssen zwei Tatsachen berücksichtigt werden: 1) Der Altersdurchschnitt der Zweitgeborenen war zur Zeit der Prüfung 6 bis 18 Monate geringer als der der Erstgeborenen; 2) die Ungleichheit der verschiedenen Skalen schliesst eine negative Korrelation zwischen IQ und CA ein, die für beide Geburtsgruppen typisch von der Ordnung 0,3 ist. Die Korrektion des Altersfaktoren, mittels der Regressionsformel, führt zur Schlussfolgerung, dass die Intelligenz keine Funktion der Geburtsordnung ist, jedenfalls soweit dies die beiden ersten Geburtsgruppen betrifft. Dies wird durch eine Studie der Erst- und Zweitgeborenen in getrennten Familien bestätigt, die in derselben Schulbevölkerung vorgenommen wurde. Nachdem man die notwendigen Korrekturen vorgenommen hatte, ergaben sich für beide Geburtsgruppen im Wesentlichen dieselben IQ'n. Dies wird ferner durch eine Untersuchung bestätigt, die die Zugehörigkeit von 427 Paaren Erst- und Zweitgeborener zu den Schulklassen zum Gegenstand hatte: es zeigten sich keinerlei Unterschiede hinsichtlich der Geschwindigkeit des Fortkommens in der Schule.

Diese Ergebnisse werden von gewissen möglichen störenden Variabeln als unabhängig erwiesen; wie beispielsweise das Alter der Mütter, Zufälligkeiten vom der Geburt, verschiedene Geschlechtsverhältnisse der zwei Geburtsgruppen und selektive Faktoren, die durch Variable der Familiengrösse bedingt sind.

HSIAO

$7.00 per volume
Single numbers $2.00
This issue, $[illegible]

MONTHLY
Two volumes per year

[illegible]
Volume [illegible], No. [illegible]

GENETIC PSYCHOLOGY MONOGRAPHS

Child Behavior, Animal Behavior, and Comparative Psychology

EDITED BY
CARL MURCHISON

Worcester, Massachusetts

Entered as second-class matter December 1, 1925, at the post-office at Worcester, Mass., under Act of March 3, 1879

$7.00 per volume
Single numbers $2.00
This issue $3.00

MONTHLY
Two volumes per year

Mar.-Apr., 1931
Volume IX, Nos. 3-4

GENETIC PSYCHOLOGY MONOGRAPHS

Child Behavior, Animal Behavior, and Comparative Psychology

AN EXPERIMENTAL STUDY OF BRIGHT, AVERAGE, AND DULL CHILDREN AT THE FOUR-YEAR MENTAL LEVEL*

From the Department of Psychology, Stanford University

By

HELEN P. DAVIDSON

*Accepted for publication by Lewis M. Terman of the Editorial Board and received in the Editorial Office, May 28, 1930.

Entered as second-class matter December 1, 1925, at the post-office at Worcester, Mass., under Act of March 3, 1879

ACKNOWLEDGMENTS

The writer wishes to express her great indebtedness to Dr. L. M. Terman under whom this experiment was carried out; to Dr. W. R. Miles for his assistance in making photographic records; to Mr. A. C. Barker, Superintendent of the Palo Alto schools, for his courtesy in permitting the use of the public schools for this experiment; to the principals of the schools in which the experiment was located; to the parents of the subjects, without whose loyal cooperation the experiment could not have been carried out successfully; and, finally, to the youthful subjects themselves, whose charming enthusiasm was largely responsible for bringing the experiment to a happy conclusion.

HELEN P. DAVIDSON

DEPARTMENT OF CHILD STUDY
ROCHESTER, NEW YORK

CONTENTS

I

HISTORICAL SURVEY AND THE PROBLEMS INVOLVED

The present study was undertaken with two objects in view. In the first place, it was desired to find out to what extent children with a mental age of four years could learn to read; and, in the second place, whether bright, average, and dull children, all of this mental age, would learn to read equally well under the same experimental conditions. A brief consideration of these two problems and the questions they inevitably raise will reveal the desirability of attempting to answer them by experimental means.

A. Can Children with a Mental Age of Four Years Learn to Read?

Biographical literature contains numerous examples of individuals who learned to read at a very early age. Before conclusions can be drawn from such data, however, two things must be borne in mind. In the first place, the subsequent achievement of these individuals indicates that they possessed superior mental endowment. Such being the case, their mental ages were doubtless in excess of their chronological ages at the time of learning to read. In the second place, the term "learned to read" may mean anything from ability to recognize a few isolated words to that of being able to read fluently simple, connected prose. Accordingly, until the mental ages of the individuals under discussion are scientifically determined, and the amount and

difficulty of the material read are definitely stated, considerable caution must be exercised in the interpretation of such data.

A recent study carried out at Stanford University by Catharine Cox (20) provides a means of arriving at a more accurate evaluation of this type of data not heretofore possible. Dr. Cox, with the assistance of two psychologists, attempted to estimate the childhood intelligence quotient for each of 300 men of genius. Two IQ's were calculated for each individual. The first one was based on all the data available for the first seventeen years of the individual's life, and the second on those achievements made between his seventeenth and twenty-sixth years. They represent the estimated lowest possible quotient that could account for the recorded facts regarding the behavior and accomplishments of the subjects during the period indicated.

The present writer selected from the abbreviated biographies appended to the study the names of all those who were definitely stated to have learned to read before five years of age. The names of those who, in all probability, did learn to read at this early age, judging from the kind and amount of reading they did when six or seven years old, but whose age at learning was not clearly stated, were excluded. There were left only 29 names. By making use of the first estimated IQ—which was almost invariably the lesser one—the mental age at which these individuals were stated to have learned to read was estimated. These names, with IQ and MA in parentheses after them, are given below.

JOHNATHAN SWIFT (125; 3-9).[1] At the age of three he could read any chapter in the Bible. He had learned to spell by the age of three or four.

ALEXANDER DUMAS (140; 5-7). Learned to read when he was four "through a curiosity to discover the history, customs and instincts of the animals whose pictures he saw in Buffon's Natural History."

THOMAS HOBBES (140; 5-7). Attended school from the age of four. Was learning Latin and Greek at six years.

ALEXANDER VON HUMBOLT (140; 5-7). Learned to read and write at four years.

CHARLES DICKENS (145; 5-7). "Charles' mother began when he was four to teach her son English and the rudiments of Latin, awakening his passion for reading and knowledge."

RALPH W. EMERSON (145; 4-4). "A week before young Emerson's third birthday his father recorded that he 'did not read very well,' but the lad progressed in favor from that time."

FREDERICK W. VON HUMBOLDT (145; 4-4). "Campe taught the three year old to read and write and within the following year instructed him in history and geography as well."

DANIEL O'CONNELL (145; 5-9). "At the age of four Daniel learned his alphabet in an hour from an itinerant 'hedge-school master.' "

LUDWIG TIECK (155; 6-2). "Ludwig could read when he was scarcely four.... When he was about four, neighbors and relatives seeing him on his little footstool eagerly reading in the Bible would shake their heads doubtfully over his unusual talent, or even refuse to believe that he was actually reading."

FREDERICK A. WOLF (155; 6-7). "When barely two, Frederick was taught to enunciate clearly, and as a game he learned sentences and verbs, first in German and then in Latin. At three he began to write and to study music. Before he was five his father had instructed him in reading aloud, thinking without writing, avoidance of ambiguity, and mental computation. At six he had made some progress in Latin and knew a little French and Greek."

LOUIS ALFRED DE MUSSET (160; 4-9). "At the age of three, Alfred began lessons in reading and writing, subjects which he pursued with great ardor in order to gain sufficient skill to carry on a correspondence with a beloved cousin."

---

[1]The figures 3-9 mean a mental age of 3 years, 9 months, and so on.

TORQUATO TASSO (160; 4-9). "Reported to have begun the study of grammar at three, Torquato was instructed in that year or a little later, by an old priest learned in the rudiments of letters and the humanities."

CHRISTOPH WIELAND (160; 4-9). "The boy's first teacher was his father who taught him to read and write when he was three, began Latin with him some time before he was six, and also developed in him a love of Science."

COMPTE DE MIRABEAU (165; 7-5). "At four and one-half Mirabeau's education began under a tutor. The boy progressed rapidly."

COUNT DE CAVOUR (150; 6-0). "At four Camillo was taught by his mother to read, but although he soon learned to read and write with ease, the passion for study did not appear until later."

LONGFELLOW (150; 6-0). Attended a private school in his fourth and fifth years where he learned his letters.

FRIEDRICH ERNST SCHLEIERMACHER (150; 6-0). "Before he had reached the age of four young Fritz had learned to read."

WALTER SCOTT (150; 4-6). "At three or four Scott learned to read in a dame's school with occasional supplementary lessons from an aunt."

GEORGE CUVIER (155; 6-2). "She . . . . taught him to read fluently at the age of four years."

BARTHOLD G. NIEBUHR (165; 6-7). "Niebuhr began in his fourth and fifth years to learn the three R's from a tutor."

VOLTAIRE (170; 5-1). "Instructed in a desultory way by the Abbé Châteauneuf, little Arouet (aged three) learned to read from La Fontaine's Fables."

SAMUEL T. COLERIDGE (175; 3-6 or 5-3). "Coleridge entered Reading School at two or three and remained in attendance there until at six. . . . Before he was five, Coleridge read the Arabian Nights, after which he was haunted by specters."

ALBRECHT VON HALLER (175; 5-3). "At three and four his greatest delight was to make books. In some of these he kept careful financial accounts; in others he wrote down all the new words he learned from day to day until he had completed a number of lexicons."

JEREMY BENTHAM (180; 5-4). "Jeremy learned the alphabet before he could talk. When he was three his father bought a Latin Grammar and other books to begin his classical education. The Greek alphabet he learned at his father's knee. . . . Fragments of Greek and Latin have been preserved written by Jeremy at four and five. . . . When only three he read with satisfied absorption Rapin's History of England."

MACAULEY (180; 5-4). "At three Macauley did not care for boys but was fond of taking his walk and telling interminable stories from his reading or from his imagination. . . . At three young Thomas read incessantly and his memory retained without effort the exact phraseology of the book."

GOETHE (185; 5-6). "From the age of three until he was six, Goethe attended a day nursery or kindergarten, and here, according to tradition, he learned to read. . . . An ABC book was purchased for young Goethe (aged 4½) and a year later a catechism with Biblical quotations was presented to him."

JOHN S. MILL (190; 5-8). "He began to learn Greek at three, and from then to his ninth year he studied Greek classics. . . . Mill read Greek and history from his fourth year onward; Plato at seven. . . ."

HUGH MILLER (135; 5-4). "At the age of four or five Hugh learned to read at a dame's school."

LEON GAMBETTA (140; 5-7). "At four years Gambetta first attended school and there he learned to read."

It will be seen that, of the above 29 individuals, only eight learned to read before they had attained an estimated mental age of five years when calculated as described, and only five at a mental age of four and one-half years or less. Two learned to read with an estimated mental age of less than four years.

These figures, of course, are only approximate, and depend for their validity on the number and accuracy of the facts given, as previously stated. They seem to indicate, however, that though a mental age of four years may be adequate for learning to read, it cannot be definitely proved from such data, and it certainly is not as common as a perusal of such biographies might at first lead one to expect.

Another Stanford study by Terman and others (71) presents some additional data of a similar nature. This study concerns gifted children (above 140 IQ) of the present decade. The parents of the children

were given information blanks in which they were asked to write answers to these questions: "Did child learn to read before starting school? At what age? How (amount and kind of help)?"

Since these data were collected while the desired information was still fairly fresh in the memory of the parents, they may be considered as probably more reliable than those of Cox, the latter having been obtained from biographies and autobiographies written usually some years after the childhood of the individual in question.

There were 552 responses to the above questions, but the more or less exact age of learning was reported for only 246 cases.

After commenting on the danger of using such insufficient data as these for statistical purposes, Terman states: "We can say, however, that at least 113 (20.5 per cent) of the 552 children learned to read before the age of 5; 34 (6.1 per cent) before 4; and 9 (1.6 per cent) before 3." When it is borne in mind that the IQ's of these subjects range from 140 to 200, it will be at once seen that these percentages would be considerably reduced were they expressed in terms of mental age instead of chronological age. Here again the evidence is that learning to read at a mental age of four years is not so common as has sometimes been believed.

Only two experimental studies seem to be available. One of these (14) is an account of the remarkable achievements of an infant who was educated by her father, a man unskilled in scientific technique. When this child, Martha, was 14 months old, her father began to teach her to read. This training was somewhat

irregular until she was 19 months old, but from then until she was 26 months old it was systematic. She was first taught to recognize capital letters, then the small letters, then isolated words, and, finally, to read sentences. Motivation was gained through games and rewards. The material used consisted of large cardboard letters, illustrated nursery rhyme and alphabet books, home-made picture books with names printed in large letters, home-made primers, school primers, and flash cards. When Martha was given her picture books to look at she had to name the word on one page before she was allowed to see the picture on the next page. She was taught no phonics, new words being simply pronounced for her. All words, except proper names, were presented to her with small letters during the pre-primer period. It is not clearly stated just how much time was spent on instruction, but, when Martha was 23 months old, she "ordinarily spent as much as an hour a day on the floor, playing with the boards (on which her reading lessons were pasted for a time), and reading the lessons aloud to herself."

The following extracts briefly summarize her achievement at different ages. At 19 months of age she could recognize and name all the capital letters; at 20 months she could recognize and name in addition all the small letters; at 21 months she had a reading vocabulary of 35 words; at 23 months she derived "mental pleasure" from reading; at 24 months she had a reading vocabulary of over 200 words; at 26½ months she had a reading vocabulary of over 700 words. By this age she had read four and one-half school primers. When 26 months old she read 44 pages (2464 words) of Gor-

don's Reader, Book One, in three and one-half hours. This contained 51 new words which were named for her. At the end of 30 hours she recognized 38 and failed in 13 of the new words.

When she was 3 years and 11 months old, Martha was given a Stanford Binet Test. By this she was shown to have a mental age of 5-6 and an IQ of 140. On this basis, assuming the IQ to be relatively constant, she had:

an MA of 1-8 when 14 months of age;
an MA of 2-4 when 19 months of age, and
an MA of 3-0 when 26 months of age.

Here, then, is concrete evidence of a child who could read primers and first readers with much fluency and expression at a mental age of three years.

That she has suffered no harmful effects from such early training there is abundance of proof. Reports of her progress, trait ratings, and other data, are obtained at regular intervals from her parents and school teachers, and kept on file at Stanford University. It may be of interest to make a few quotations from these.

On January 4, 1922, when Martha was not quite six years of age, she was given a second Stanford Binet, according to which she had an MA of 8-9 and an IQ of 142, the latter being two points higher than that obtained on her first test. She passed the reading test in year ten, as also the "absurdities" test, but failed in all tests of the twelve-year level.

From January 4 to March 4, 1922 (at the age of 6-0), she read 21 books, chiefly the Burgess books, according to the Reading Record kept for that period. According to the Home Information Blank filled out

at this time she was in the habit of reading 20 hours weekly before she was five years of age, and the same number of hours when between five and six years of age. Despite this, she was very normal in her play interests, and enjoyed playing out of doors with other children.

June, 1923 (age 7-5). Martha is in the third grade. According to the Stanford Achievement Test given then she had an educational age of 11-1. By it her reading is equivalent to low sixth grade, her arithmetic to high fourth, and her spelling to low fifth. Her teacher reports her as interested in her school work, popular with other children, and not at all conceited.

November, 1927 (age 11-10). Martha is now in high eighth. The field worker reports her as a big girl for eleven, friendly and communicative, and free from self-consciousness. She expressed the desire to become a writer, and has already written "two little books" of the Burgess type. Her teacher reports her at this time to be a normal child, apparently in good physical condition, and possessing "splendid poise." She states further that Martha shows special interest in literature (reading and composition), evidence of a good imagination and "a definite and intelligent enjoyment of good literature." She has no weak subjects, is serious and ambitious, and responds well to discipline.

A perusal of these reports cannot fail to convince one that Martha's early intensive training has failed to produce any abnormalities, whether physical, mental or social. She appears to possess a superior physical and mental endowment, but her development is all-

round and well balanced. She displays normal child-like interest in play activities and companions, and, thanks to good home environment, she is free from self-consciousness and conceit. The early "mental pleasure" in reading which she evinced at the tender age of two years has apparently never abated and is now bearing fruit. Her further progress will be followed with much interest.

The second study was carried out at Stanford University by Muriel Brown (19). It consists of two distinct experiments. In the first experiment Miss Brown undertook to teach a young girl (CA = 3-6; MA = 3-6) and a boy (CA = 3-6; MA = 4-10) to read; in the second, she studied word learning with a group of 13 children aged 22 months to 5 years and 11 months. The first experiment is of the greater importance here and will be reviewed. The two children were taught individually in their homes, five days a week for a period of three months. "Each lesson lasted from 20 to 30 minutes and included a number of games apart from the reading instruction." The materials used consisted of posters, flash cards, and stencils. The posters were reproductions of the pages in the Everyday Classics Primer. "The method of teaching was essentially the same in the case of both children; a pre-primer period of word learning, motivated by games, being followed by poster and primer work on the story of the Three Bears." Miss Brown did not give any reading tests, but from two word-learning curves, which she states are only approximate, it is seen that Catherine could recognize 34 words and Roger 27 words at the end of the experiment—presumably out

of context. At the end of the experiment the author states that "Catherine, in 36 lessons, learned to read 11 pages of her primer. Roger in 26 lessons, reached the same point. Both children could read their charts, and the re-combination of words upon the charts, with perfect ease."

The distinctive feature of her method was the use of stencils. A printed word, together with the object it represented, was shown to the child. The word was named for him, and he was then given a cardboard stencil of the word which he traced ("redded") on a piece of paper. When the stencil was removed the printed word was found on the paper. After having learned several isolated words in this way, the child began to study the poster of the Three Bears, and, later, the same story in the primer. New words were learned by means of the stencil.

Miss Brown attributes her success to:

(*a*) "The emphasis placed upon the kinaesthetic factor in learning, through the stencils."

(*b*) "The 'objectivity' of the didactic material employed."

(*c*) "The variety of the devices used to secure spontaneous attention."

The present writer is inclined to believe that Miss Brown overestimates the importance of the kinaesthetic factor in her method. The value of the stencils lay, rather, in their interest as a pleasurable activity, which, at the same time, drew attention to the form of the word.

As a result of her two studies Miss Brown concludes that:

1. "*Word learning* is possible at a mental age of less than two years."

2. "*Word-relating,* reading proper, involves reasoning processes which are certainly operative at three, if not before, but cannot function without certain raw materials, ideas to be expressed, and a store, however meager, of mastered words to which experience has given meaning."

3. "Striking individual differences in temperament, training and environment determine, for each individual child, the age at which formal instruction should begin."

4. "The most effective method which can be employed is the method which best suits the needs and capacity of the child to be taught, but, in all cases, is probably one which emphasizes the kinaesthetic factor."

These two studies demonstrate the success of two subjects, Martha and Catherine, in learning to read at a mental age of less than four years, and of one subject, Roger, at a mental age of less than five years. They do not show whether this result can be regularly accomplished. Should the present study show that a mental age of four years is adequate for learning to read, of what value will it be? The following section attempts to answer this question.

The mental age requisite for learning to read is a matter of great practical importance. Despite the fact that the reading situation is an artificial one, in the sense that printed symbols were invented by man, yet the physiological apparatus and the psychological processes involved are sufficiently mature at the mental

age of four years that they may be said merely to await the proper stimulation to bring them into operation. (In the case of Martha they were present at the mental age below three years.) This would seem to suggest that ability to perceive printed symbols and to attach meaning to them are simply parts of, or stages in, the development of perception and of meaning, a development that has been going on since birth.

The practical aspects of the problem are numerous and important.

1. The study should lead to an analysis of first-grade reading as at present taught in an effort to locate the cause of the high percentage of failure that is found there, and to test out the current belief that a mental age of six years is necessary to achieve success in first-grade work. The seriousness of these problems is shown by the fact that "In the average city approximately a fourth of the pupils fail in promotion at the end of the first year" (10), and that "children who are below six years in mentality find it very difficult to do work that is satisfactory in the first grade, and are more likely to do inferior or very inferior work, while most children who test above six show accomplishment that is of average or superior quality" (23). The chief emphasis in the first grade is placed on reading, hence failure to do first-grade work is practically synonymous with failure to learn to read. H. I. Baker (1) reports from 30 to 35 per cent of first-grade children failing each year, and attributes the majority of such failures to mental ages too low to enable the children to learn to read. Reed (66), in a recent study of first-grade promotion, states that "The evidence that read-

ing is the decisive factor is further sustained in the answers of 166 administrators to the question, 'Would a child with a satisfactory rating in reading be denied promotion because of a low rating in any other factor?' Of these, 119, or 71.6 per cent, replied that a child with a satisfactory rating in reading would not be denied promotion even though he had a low rating in other factors; 2, or 1.2 per cent, replied 'Yes and no'; 45, or 27.1 per cent, replied that he would be denied promotion." Terman (72) says that results "indicate that at every step, and in every domain, the mental level which has been attained sets rather hard and fast limits to the possibility of training. Children below the mental age of five years do not learn to read in the public schools. It can not be said that these feats of learning are impossible at the respective levels mentioned; all we know is that they do not ordinarily occur under present day educational treatment."

2. In the second place, there is the question of the amount of time spent in teaching reading in the first grade. Holmes (49) reports that in 50 representative American cities, an average of 266 hours per year, or 30.6 per cent of the total recitation time, is devoted to reading in the first grade, the lowest amount of time being 100 hours, and the highest 570. This means that, on the average, 412 minutes per week, or 82.4 minutes per day, are spent in teaching first-grade reading. In a Stanford study, Dorothy Wilson (79) found that from 50 to 115 minutes per day were spent in teaching reading in the first grades of the six schools in Palo Alto for the year 1924-1925. A survey of California courses of study indicates that 100 minutes per day is

considered a very usual amount of time to spend in teaching primary reading. The *School Manual for Santa Clara County* for the year 1927 suggests that a minimum of 525 minutes per week, or 52 per cent of the total teaching time, be devoted to the teaching of first-grade reading, while the *Manual and Course of Study for Los Angeles* for the year 1927-1928 suggests a minimum of 100 minutes per day, or 42 per cent of the total teaching time, for schools having a 240-minute day. The above is indeed evidence that reading is of paramount importance in the first grade.

3. In the third place, not only in amount of time, but also in actual money spent, reading receives by far the largest share compared with all other subjects in the elementary school curriculum. In a recent book by O'Brien (61), the following statement is found: "Fleming, in an unpublished study, found that reading consumed approximately 2/3 of the total expenditure in the first grade, while its average cost per grade for the first six grades was more than the combined cost of any other two subjects." He gives the following table, taken from Fleming (27).

TABLE 1

THE PORTION OF EACH THOUSAND DOLLARS SPENT FOR THE TEACHING OF READING IN EACH OF THE FIRST SIX GRADES AS COMPARED WITH ITS NEAREST COMPETITORS

(From Fleming)

Subjects	First grade	Second grade	Third grade	Fourth grade	Fifth grade	Sixth grade	Average per grade
Reading	611	407	307	240	150	156	312
Arithmetic	5	101	176	187	181	190	140
Language	95	110	126	130	178	105	124
Music	86	90	84	67	58	67	75
Spelling	3	92	90	93	80	71	71
Geography			9	102	124	152	64

4. The possible true cause of failure in first-grade reading must next be considered. The above quotations, setting forth the high percentage of failure in first-grade reading despite the large amount of time and money spent on it, reveal the existence of a very real and serious problem. If the present study should prove that it is possible to learn to read at a mental age of four years, in contradiction to the general belief that a mental age of six years is necessary, to what can the above failure be ascribed? It is beyond the scope of the present study to answer this question, yet it may not be out of place to call attention to one possible source of difficulty, namely that of teaching method. Despite the fact that reading has been taught for two thousand years and more, it must be confessed that the best method of teaching it is largely a matter of conjecture and opinion, of trial and error. It has remained for the present century to make even the beginnings of a scientific inquiry as to how children really learn to read. Gates (38), too, suggests that in all probability a better teaching technique may solve the problem of first-grade failure. He says: "The question is whether reading is intrinsically so difficult a function that many normal children are unable to master it until after the sixth year. Theoretically, there is no reason to believe that it is more difficult to associate things and events with visual symbols than with auditory signs. Indeed, some evidence is being gathered which indicates that very young children can, without marked difficulty, learn to comprehend printed words." He goes on to point out that he is not advocating that reading should be taught at three or five or six years of age, but is

"questioning the validity of the implication that learning to read is fundamentally and necessarily difficult and late in appearing and whether many first-grade children are intrinsically unable to learn to read. Is it not possible that the fact that they often do not succeed in the first year of school is itself a suggestion that our methods of teaching afford too little exact guidance? Perhaps the difficulty is that the child is bewildered by the complexity of tasks presented when reading for thought-getting is employed from the first, with little real direction, and that too little vital help is secured through transfer from the supplementary practice in phonetics and the like."

5. Finally, there is the question of the optimum age to teach reading. To prove that it is possible to learn to read at a mental age of four years does not necessarily imply, however, that that is the best age to do so. Perhaps reading could be more economically and pleasurably learned at five or seven or nine years of age. Further experimentation is needed before this question can be satisfactorily answered. Nevertheless, it is of interest to note in passing some of the diversity of opinion in regard to this matter.

Two experimental studies attempt to throw light on this problem. Arthur (15) found that a mental age of from 5½ to 6 years was the optimum age for teaching sight words, 6½ to 7 for phonetics, and 6½ to 7 for comprehension. She states: "Efforts toward reading comprehension for pupils with a mental age of less than 5½ years were largely wasted. A mental age of 6 to 6½ was in general necessary for standard first grade achievement." In England, where the compul-

sory age for beginning school is five years, it being optional at three years, Winch (80) draws the following conclusions from a study which he conducted: "English children, under our conditions of school organization, may quite profitably begin to learn to read at an average of 5 years 3 months."

Amongst others who express an opinion as to when reading should be taught are the following. Judd (50) and Wheat (78) believe that "the age of six is the crucial age." Klapper (51) thinks it should be taught at an age as early as the child will allow. Experience, he says, shows this to be about the age of seven. Stone says (69): "By the age of six most children have reached the stage where they desire to learn to read for themselves." Dewey (21) states that, "While there are undoubted exceptions, present physiological knowledge points to the age of about eight years as early enough for anything more than an incidental attention to visual and written language forms." Patrick (63) would defer it to ten years or later, saying that for the years from five to ten the "let alone" plan of education is of the most value, and in support of this contention, he says, "We can understand, therefore, that it will demand a considerable maturity in the child before he is ready for that which has developed so late in the history of the race. The language of the child, like that of primitive man, is the language of the ear and tongue. The child is a talking and hearing animal."

Among such diversity of opinion experimental evidence alone can settle the issue.

B. Will Bright, Average, and Dull Children, All Possessing the Same Mental Age, Learn to Read Equally Well under the Same Experimental Conditions?

The question is whether the same mental age always means equal capacity to learn, irrespective of the chronological age. This is another problem to which no satisfactory answer can be given as yet, though it has received some consideration in the last two or three years. Four of the twelve studies which deal directly with this problem have been published since the present experiment was begun. The studies will be taken up in chronological order and the chief emphasis will be placed on reading ability when this has been investigated.

Colvin (3) gives a very brief account of a study in which five normal children were compared in ability to cancel *a*'s with five subnormal children with whom they had been matched by Binet mental age. In every case the normal children made greater improvement than the subnormal children and showed fewer fluctuations.

Woodrow (13) used four groups of subjects, normal practice and control groups, and feebleminded practice and control groups. Their average mental age was 9 and their chronological age below 14 years. Besides being comparable as to mental age, they had been selected also on the basis of equal ability in sorting sticks according to length, sorting colored pegs, cancelling letters, and cancelling geometric forms. The normal and feebleminded practice groups were then

given two practices a day for 13 days in sorting gun wads according to five kinds of geometric forms which were marked on them. At the end of that time both groups showed 45 per cent improvement. As a result of this experiment Woodrow concludes that feeble-minded children not only can do the same things that normal children of the same mental level can do, but can learn them equally well.

Murdoch (8) comes to the opposite conclusion of that of Woodrow as a result of an experiment which she carried out with feebleminded children. In 1917 she gave a number of achievement tests in reading, spelling, arithmetic, etc., to 21 feebleminded subjects whose chronological ages ranged from 12 to 23 years and mental ages from 7-11 to 11-1. She repeated these tests one year later. As she had no control group, she compared the 1917 results with the norms for third grade, and the 1918 results with the norms for fourth grade, her justification for doing so being that the average mental age of her group in 1917 was 9-2, and hence comparable to Grade Three. She draws the following conclusions from her data. (*a*) "In every case of the school subjects in which comparison can be made between feeble-minded and normal, it will be seen, that our group of feeble-minded in 1917 did better than normal children of the same mental age." (*b*) "The Thorndike Reading Scale (row 10) shows that the feeble-minded in 1918 did better in this test than normal children of the same mental age (4th grade)." (*c*) "The rate of learning of feeble-minded children in these elementary school subjects is less than the rate of normal children of the same mental level." (*d*) "As

for the so-called mental age, whatever this well known term may mean, according to the results of the present study it does *not* mean either ability to do the same grade of work as done by others of the same mental age, in most of the elementary school subjects; nor does it mean ability to make the same rate of progress in these same school subjects."

Owing to the small number of the subjects, the wide range of ages, and the criteria used, the results of this study cannot be considered conclusive. However, judging from the author's own statement that the feebleminded actually did better than the normal subjects with whom she compared them in 1917, it would seem that mental age for the two groups must have something in common. The subsequent inferiority of the feebleminded group one calendar year later seems, in all probability, to be due to a slower rate of mental growth rather than to a qualitative difference in mental age in the strict sense of the word.

C. T. Jones (6) made an intensive study of five very bright and five very dull boys of the same mental age, that of ten years. He found certain qualitative differences between the groups. The young bright boys equalled the feebleminded boys in speed of reading and excelled them in comprehension of reading material. He disagrees with Woodrow's conclusions concerning the learning ability of the feebleminded, basing his own conclusions on the result of the learning test (Woodworth-Wells Substitution Test), which the very bright learned to do more quickly and accurately than the feebleminded, but he believes that this ques-

tion cannot be settled conclusively without further experimentation.

The next study to be considered here was undertaken with a somewhat different aim in view from that under discussion, but it has some contributions to offer in the general field. Chapman and Dale (2) selected two groups, a Young Bright and an Old Dull group, from 5000 subjects who had been given the N. I. Tests. There were 50 subjects in each group, paired according to N. I. T. score, with scores ranging from 70 to 119. The Young Bright were less than 10 years old and the Old Dull more than 13 years of age. The individual tests were then examined to determine which were measures of native endowment and which of environment. As a result of their analysis, these writers make the following statement which is pertinent to the present topic. "If, as we are bound to assume when we employ a group test, the same total represents the mental age, whatever may be the chronological age, this study shows that the mentality of the Young Bright is different from that found in the Old Dull. It also shows that intellectual development is not marked by a simultaneous and uniform improvement in all types of mental function. The Old Dull of mental age x is the superior, equal or inferior of the Young Bright of mental age x, mental age being measured by totals, according to the type of test used."

Taylor (9) gave superior, normal, and feebleminded subjects of varying chronological ages, but with mental ages of six and seven years, various types of tests, such as the Army Alpha, Pintner Mental Survey, Thorndike Non-Verbal, etc. Though she, like Jones and

Chapman and Dale, found differences among the groups, she comes to the following conclusion: "It may be concluded that there is a greater similarity than difference in the responses of individuals of six and seven years mental age from four to twenty years chronological age."

Merrill (7) made a careful study of the achievement of mentally retarded children. In one part of her study she compared a group of 52 mentally retarded children, whose chronological ages ranged from 7 to 18 years, and whose average IQ was 69.8, with a group of 52 normal children with whom they were matched mental age for mental age. The average IQ of the latter group was approximately 101. Both groups were given numerous achievement tests, including three reading tests. On examining the results of the reading tests, she found that the normal excelled the retarded at every mental age level except mental age seven, and, in one test, at mental age eight. After a careful statistical analysis of the results, Dr. Merrill states: "The experimental evidence is conclusive that the retarded children show no marked difference in ability in reading, arithmetic and spelling. The slight superiority in arithmetic and spelling indicated in mean scores is too small to be significant." And, further, "In general, the results of the investigation indicate that there are characteristic differences in mental traits between normal and retarded children of the same mental level but that 'capacity for learning' in the two groups is not significantly different at any level."

Torgerson and Shuman (11) compared bright, average, and dull children, all with a mental age of 12

years, in the Thorndike-McCall Reading Tests. The bright group gained a reading quotient of 111.1, the average and the dull one of 104.4. The writers conclude that the higher the IQ the higher is the score in reading. However, the statistical significance of the differences found appears questionable. These same investigators next compared the Stanford Achievement scores of 112 subjects in the fifth and sixth grades who had mental ages from 11 to 11-11. They found that the normal group had the highest accomplishment quotient and the bright the lowest. The bright and normal groups gained a Reading Quotient of 110.0, the dull one of 107.5. These writers believe their data indicate "that the accomplishment of pupils of the same mental age who are given the same opportunity for learning will, in general, vary in direct ratio with the I.Q."

F. T. Wilson (12) gave several group tests, including the Stanford Achievement Test, to 2000 subjects in grades 3A to 8A inclusive. The subjects were then divided into three IQ groups, a high, a normal, and a low. At each mental age at which all three groups were represented, namely, mental ages 10 to 13 inclusive, the low-IQ group had the highest score 12 times, the normal group 3 times, and the high group only once. At these same four levels, the low group obtained the highest score in reading twice and the normal group twice.

Cunningham (4) compared the Thorndike CAVD test results of normal preschool children with those of adult imbeciles, matched in mental age. After a careful statistical analysis of the scores made by the two

groups, he concludes that there are characteristic differences between them: that certain tests are easier for the imbeciles because of their greater experience, and some for the children because of their greater intelligence. Despite these facts he believes it permissible to use the same tests on dissimilar chronological age groups, provided they do not differ too much in this respect. This would seem to indicate that, like Taylor, he believed that the similarities outweigh the differences.

The last, and most recent study, is one by Fox (5), in which 63 feebleminded children are compared with 63 normal children of approximately the same mental age. The range of mental age for the feebleminded group was 4-6 to 9-9, with a median MA of 7-3; while that for the normal group was 5-7 to 9-9, with a median MA of 7-4. The range of IQ for these groups was 33 to 83, with a mean of 56 for the former; and 76 to 116, with a mean of 83 for the latter. Both groups were given four primary group tests. Fox then compared the two groups with regard to their successes in the different items. Like Cunningham, Fox found that the feebleminded excelled in tests involving experience. Somewhat more pertinent to our problem is the following conclusion: "The defectives are not able to read and understand the meaning of words to the extent the normals do. In the Word-Comparison Test they fall far below their mental age level."

The above summary of the literature dealing with comparisons of bright and dull children of the same mental level reveals a fairly general belief that there are characteristic differences between the groups, but

what these differences are and how significant they are is not so clear. Five of the studies discussed deal with achievement tests. Torgerson and Shuman found that the bright subjects did better than the dull ones; Murdoch, in her 1917 results, and Wilson found that the dull did better than the bright; and both Jones and Merrill obtained varying results, but the latter found that the differences between the groups were not statistically significant.

Three studies, those by Chapman and Dale, Cunningham, and Fox, analyze mental-test items. (Taylor's study might also be included here.) All found characteristic differences between the groups. Chapman and Dale are emphatic that mental age does not mean the same thing for the two groups; Cunningham and Taylor believe the similarities outweigh the differences; Fox takes no definite stand on this point. Of the authors of the two remaining studies dealing with various types of material, Woodrow definitely states that mental age does mean the same thing for both groups, and Colvin found the bright superior to the dull.

Six of the above studies contrast the reading ability of bright and dull subjects. Torgerson and Shuman and Fox found that the bright read better than the dull; Wilson and Murdoch that the dull read better than the bright; Jones that the bright equalled or excelled the dull, depending on the phase of reading measured; Merrill that the dull excelled the bright in the early mental levels but that the reverse held true in the higher levels. Merrill was the only investigator to attempt to establish the statistical significance of the

differences between the groups and found it to be very slight.

Attention should be called to the following considerations: (*a*) If the present study should prove that bright, average, and dull children, all with the same mental age, do not learn equally well under the same experimental conditions, it cannot be immediately inferred that one group is superior in learning ability to the others. It may be that each group requires a different method of instruction, and that these special methods will at least reduce the differences in achievement between the groups, if they do not succeed entirely in abolishing them. This will necessitate a search for the best method for each group and the organization of the schools into suitable groups. (*b*) Before it can be concluded that mental age does or does not mean the same thing irrespective of chronological age, it will be necessary to search for the existence of factors other than brightness which might be responsible for the obtained likenesses or differences in achievement.

II

NATURE OF THE PRESENT EXPERIMENT

Aim

As previously stated, the aim of this investigation was twofold, namely, to find out (*a*) to what extent children with a mental age of four could learn to read, and (*b*) whether bright, average, and dull children, all of this mental age, would learn to read equally well under the same experimental conditions.

The investigation was carried out at two different periods. The part dealing with the dull and average took place from February to June, 1926; the part dealing with the bright group took place from November, 1926, to March, 1927. For convenience, since the methods with all three groups were essentially the same, the experiments with the dull, average, and bright groups will be discussed as if they took place concurrently.

Criteria for the Selection of Subjects

By selecting the mental age of four years as the mental level to be studied, it was possible to secure three groups of children with no previous experience in a reading situation. This is a very desirable condition and one that is rarely attained. It eliminated entirely the factor of unequal training in the item to be studied.

The following desiderata for the composition of the three groups were set up:

1. The mental ages of all three groups should be

between four and four and one-half years as measured by the Stanford Binet Test.

2. The chronological ages of the three groups should be as follows:

a. The normal group between 4 and 4½ years.

b. The dull group between 5 and 5½ years.

c. The bright group between 3 and 3½ years.

3. The intelligence quotients of the groups should be as follows:

a. The normal group between 90 and 110.

b. The dull group below 90.

c. The bright group above 110.

4. There should be four or five children in each group.

5. The parents of the children should be white Americans, born in this country. This was to eliminate all question of language difficulty in the home.

6. Only children who had successfully passed the four-year test involving discrimination of forms in the Stanford Binet should be selected. This is one of the easiest tests at this level and so it did not complicate the selection of children.

It proved a much more difficult task to locate suitable subjects than was at first anticipated. The following sources were appealed to: The kindergartens of the schools of Palo Alto (for dull five-year-old children), the pupils of all grades in two elementary schools for names of young sibs, City Hall birth records, school nurses' reports, city householders' canvass, a university club, neighbors, etc.

After a month's laborious search it was found necessary to extend the mental-age range in order to get the

quota of subjects. Accordingly, a subject with a mental age of 3-8 and one with a mental age of 3-10 were added to each of the average and dull groups. When selecting the bright group, some months later, an effort was made to match them mental age for mental age with the former groups.

Table 2 gives the Stanford Binet Test results in detail. It will be noted that one of the normal group had to drop the experiment after six weeks' work because of illness. This left four subjects in each of the dull and average groups and five (including a pair of twins) in the bright group. Two of the five-year-old children were already attending kindergarten. One, Donald, had been attending it for several weeks, and the other, Bell, for only two weeks. None of the other children had had any kind of school experience.

In every instance, all testing, teaching, and interviewing of parents were done by the writer.

General Plan of the Experiment

It had been decided to have the children meet in groups in a central meeting place rather than to take them individually in their homes for the following reasons: The home environment would have differed for each child; the close proximity of his mother and sibs would doubtless have proved a very distracting factor for the subject; it would have been very difficult to transport and use in the home the material planned for the experiment; it would have been very time-consuming for the experimenter to go from home to home; it was thought that group competition might prove to be a useful factor in the experiment.

The four- and five-year-old children had been located within a single school district. Permission to use a vacant room in the neighboring school was readily granted by Mr. A. C. Barker, Superintendent of Schools. The three-year-old children, whose homes were scattered over a wide area, met in another school.

TABLE 2

STANFORD BINET SCORES AT THE BEGINNING OF THE EXPERIMENT

Group	Names		CA	MA	IQ
	Jean		3-3	4-6	138
	Paul		3-4	3-10	115
Bright	Peter		3-4	4-0	120
	Kate		3-3	4-2	128
	Sally		3-0	4-2	139
	$N = 5$	Av.	3-2.8	4-1.6	128
	Polly		4-0	3-10	96
	Leo		4-2	3-8	88
Average	Elsa		4-6	4-6	100
	Lela		4-4	4-6	104
	Dora		4-1	4-2	102
	$N = 5$	Av.	4-2.6	4-1.6	98
	$N = 4$	Av.	4-3.25	4-2.5	98.5
	George		5-0	3-10	77
	Edgar		5-6	4-2	76
Dull	Donald		5-1	4-2	82
	Bell		5-0	3-8	73
	$N = 4$	Av.	5-1.75	3-11.5	77

The average group met from 9 A.M. to 10:30 A.M. five days a week for four and one-half months; the dull group met from 10:30 A.M. to 12 M. during the same period. Only a short part of this time was devoted to the actual experiment. Each child received an individual 10-minute reading lesson each day; while the group, as a whole, usually received a brief group game, daily, in addition. The rest of the time was devoted to kindergarten games and play. While the individual reading lesson was in progress the other children sat

at the table, or on the play-rug if they so preferred, cutting, pasting, crayoning, stringing beads, building castles, etc. It was very necessary to allow a margin of time so that there would be no feeling of hurry. Small children are very slow both at beginning and at laying down a task. It was in order to make it unnecessary to hurry a child unduly, and to have an abundance of time to care for any emergency that might arise without interfering with the time necessary for the experiment proper, that the relatively long meeting time was decided upon.

As for the three-year-old children who had barely emerged from babyhood, it was difficult to foresee whether they could be successfully handled in a group like the other children. Since it was extremely desirable that they should be so handled in order to obtain results strictly comparable to those obtained from the older groups, it was decided to at least attempt it, the parents being warned, however, that, if it were not successful, each child would be taught in his own home. For the first week they met at school for only half an hour daily so that it might be seen how they would adjust to the novel situation. Their enjoyment was so very evident—they wanted to come on Saturday and Sunday, too—that the time was extended, and the daily routine was planned to duplicate that of the older children. It so happened that all five children could not meet at the same hour because of the occurrence of morning naps and difficulties in transportation. Accordingly, two groups were formed, one of three and one of two subjects, meeting at different hours.

Since the experiments were carried out in the public

schools, the public school holidays had to be observed. The average and dull groups met for instruction on 80 days between February 1 and June 2, 1926, inclusive. In this period occurred the Easter vacation of one week and two single-day holidays, neither of which is included in the 80 days. The testing program began on June 3 and was spread over a period of about two weeks in order to avoid fatigue. This time, also, is not included in the 80 days of instruction. The bright group met for instruction on 76 days exclusive of holidays, namely from November 8, 1926, to March 15, 1927. Its testing program began on March 16 and lasted about a week. This experiment was interrupted by two long vacations, one week at Thanksgiving and two weeks at Christmas, with one single day in February, making 16 school days of vacation in comparison with 7 for each of the other groups. The experiment with the bright group ended somewhat abruptly, and the testing program was begun a little earlier than planned, because of the fact that the sister of one of the subjects had developed measles. This subject was quarantined, but, in order not to lose her record, the experimenter visited her daily in her home, and gave her the complete set of tests. A few days later, she, too, contracted the disease.

There was slightly more irregularity of attendance among the average and dull children than among the bright. In view of the fact that the experiment ran over a period of four and one-half months and that the weather was frequently most disagreeable, it is surprising that the absences were not more numerous than they were. The two most trying cases of sickness oc-

curred near the end of the experiments. The best subject of the average group was absent through sickness during the whole of the last week of instruction and for most of the testing period. The experimenter went daily to her home and tested her there, but she was still somewhat pale and upset. The other case, in the bright group, has just been described.

Words are inadequate to describe the cheerful and unfailing cooperation of all the parents throughout the experiment, as, twice daily for a period of four and one-half months, they ungrudgingly upset their household plans to take their children to and from school. Some of the parents appreciated the experiment as a scientific study more than others, but each and every one cooperated so loyally that it was possible to bring the experiment to a successful close with the loss of only one subject, and that was due to genuine illness near the beginning of the experiment.[2]

Testing Program[3]

Complete vocabulary tests and several unstandardized reading tests were given at regular intervals during the experiment, while at its close and again some weeks later a battery of tests was given. The final testing program for Groups A and D began on June 3

---

[2]The writer wishes to express her thanks to Mrs. Howe, kindergarten teacher, for her great kindness in permitting two subjects, who had a long distance to come to school, to remain in the kindergarten room for half the morning.

[3]Throughout this study
Group B will stand for the bright three-year-old subjects,
Group A will stand for the average four-year-old subjects, and
Group D will stand for the dull five-year-old subjects.

and covered a period of almost two weeks. Through the courtesy of the kindergarten teacher, the children stayed in the kindergarten room during the whole of this period and came singly to their own room for the tests. As already stated, Elsa of Group A was tested daily in her home because of illness. The final testing program for Group B was begun on March 16. One subject, Jean, was tested in her home because she was confined by quarantine.

All groups were retested several months after the end of the experiment. The parents had not been told at the close of the experiment that retests were being contemplated, but in October of the same year, four months after the close of instruction, the tests were repeated with all but one subject of Groups A and D, that one having moved with his family to Pasadena. Two others had left town, but arrangements were made and they were duly tested by the writer. The retesting of Group B, like its final testing, was begun very hurriedly. It was accidentally found out that three of these subjects were leaving the state in a week's time, and arrangements were immediately made to have them brought daily to the campus for the retesting program, beginning May 24, 1927, nine weeks after the close of instruction. Unfortunately, Jean, who was tested under trying circumstances before, was taken ill after one day's testing and she had to be given the bulk of the tests on the day before she left town. Because of illness and the excitement at the thought of returning to her own home, she did not do herself justice. The same period of time, therefore, did not elapse between the final tests and the retests for all the groups as had

been planned. This is to be regretted since it makes the comparison of the results less satisfactory than would otherwise have been the case.

The following tests were used for the finals and retests: (*a*) a complete vocabulary test, (*b*) an improvised reading test, (*c*) Wag and Puff Reading Test, (*d*) The Merrill Preliminary Tests, (*e*) The Detroit Word Recognition Test, Form A, (*f*) The Pressey First Grade Reading Test, and (*g*) The Stanford Binet Test.

The two most advanced subjects were given two additional tests: (*a*) the re-arranged story, "Puff is Sleepy," together with a vocabulary test of words and phrases, and (*b*) an unknown story, "The Boy Who Ran Away," together with a vocabulary test.

In addition to the above tests, but at the time of the retests only, the most successful subject was given: (*a*) Haggerty Reading Test, Sigma 1, and (*b*) Gates' Primary Reading Tests. Further, Dr. W. R. Miles, Professor of Experimental Psychology at Stanford, was kind enough to make a photographic record of this subject's eye-movements while she read an unfamiliar story, viz., Merrill's Primary Reading Test, Form 2, the subject immediately afterwards answering eight questions about the story. Finally, Dr. Miles took a moving picture of this subject's eyes while she read another unfamiliar story.

Materials Used in the Experiment

These included the following: (*a*) A graded series of "Massed Forms" especially designed for the teaching of beginning reading for this experiment. (*b*)

Flash cards on which words were stamped in black type by means of a printing set, the height of whose small letters was one-half inch and of the large letters three-quarters of an inch. This will be referred to as Type No. 1. (*c*) Charts, 22 by 28 inches, on which stories were printed by means of Type No. 1. (*d*) A rubber type printing set used for printing directions for seat-work. The size of the small letters was about three-tenths of a centimeter and of the large letters about one-half of a centimeter. This will be referred to as Type No. 2. (*e*) A "reading rack" made by sewing strips of manila board 2½ inches high to a large sheet of the same, thus making shallow pockets into which cards bearing words could be slipped and sentences thus "built up." This method was used in the early lessons. (*f*) The Wag and Puff Primer by Marjorie Hardy, published by the Wheeler Publishing Company, Chicago, 1926. (*g*) A number of picture books, story books, and primers displayed on a table. The subjects were free to look at these whenever they desired, and the experimenter occasionally read from them, the object being to stimulate an interest in books. (*h*) A variety of kindergarten materials was kept within easy reach of the subjects. It was destined to keep the subjects happy and busy while the individual reading lessons were in progress, and consisted of wooden beads to string, peg boards, plasticine, plain and colored paper, pencils, crayons, scissors, paste, building and design blocks, animal stencils, old magazines, etc.

III

METHOD

During the first week the experimenter devoted most of her time to training the children to use kindergarten material in order that they might work independently with this while the individual reading lessons were in progress. At the same time the children had an opportunity of becoming acquainted with each other and with the experimenter, and of becoming adjusted to the new situation.

Considerable attention had been given in advance to the development of a teaching method. Studies too numerous to mention, bearing on every phase of the reading situation, were consulted and a method containing some novel features was finally devised. A description of the method and reasons for its adoption will be given.

What Is Reading Ability?

Ability to read seems to involve three things, viz., (*a*) ability to associate a word name with its printed symbol, (*b*) ability to discriminate between printed symbols, and (*c*) ability to derive meaning from printed symbols. The first of these implies the second, and the second seems of prime importance to beginning reading. The writer believes that perceptual discrimination is the foundation of reading. What does a child see when presented with the printed word "cat"? And how does he distinguish it from the printed word "dog"? If these two questions could be

answered, then much would have been accomplished towards the attainment of a satisfactory method. Since this problem of the perception of words is of great importance, the conclusions of experimenters in this field will be surveyed.

By means of tachistoscopic exposure of letters, words, and sentences, and through the analysis of words frequently confused, experimenters have attempted to determine how words are perceived. There is considerable difference of opinion as to which part of the word is of primary importance for its recognition, but it is evident that many factors may act as cues.

Some writers believe that words are read as wholes, and not letter by letter, basing their conclusions largely on the facts that (*a*) words can be seen at a distance too great for the individual letters composing them to be recognized, and (*b*) words can be grasped in less time than is necessary to grasp the same number of letters when they do not form a word. Other writers deny that words are seen as wholes, or, at most, they ascribe to the total form of the word a rôle of secondary importance. To them, the individual letters, in various ways, are the chief factors. Several writers believe that at one time certain parts of the word may act as cues, while at another time certain other parts of the word may so act. Table 3 briefly summarizes the chief factors used in the recognition of words, and states by whom they are so described.

It will at once be seen from Table 3 that the perception of words presents a complex situation. The majority of writers apparently believe that words are seen as wholes, but little is known of the relative import-

ance of the other factors. It may be that certain factors, such as the length or the total form of the word, serve to give a general orientation of the word, and the other factors, such as first and last letters, projecting letters and so on, supply the needed detail that leads to its complete apprehension. It is important to note,

TABLE 3

MAIN FACTORS DETERMINING THE RECOGNITION OF WORDS

Cues used in recognizing words	Name of experimenter
Words perceived as wholes	Armaignac (85), Bowden (86), Cattell (88, 90), Dearborn (95), Erdmann and Dodge (97), Hamilton (101), Huey (105), Pillsbury (112)
Determining letters call forth their sounds which then suggest the word	Goldscheider and Müller (100)
Sounds of known letters suggest the word	Winch (80)
Dominating or domineering letters determine recognition	Dearborn (95), Huey (105), Messmer (109), Zeitler (116)
Upper part of words more important than the lower part	Javal (106)
First half of word more important than the last half	Goldscheider and Müller (100), Hamilton (101), Huey (105), Pillsbury (112)
First letter of word used more often as a cue than last letter	Gates (99), Meek (111)
Last letter used more often as a cue than first letter	Bowden (86)
Last two letters used more often as a cue than first two letters	Meek (111)
Length of word is important	Bowden (86), Cattell (90), Dearborn (95), Erdmann and Dodge (97), Gates (99), Huey (105), Messmer (109), Pillsbury (112), Zeitler (116)
Trivial details in words used as cues	Bowden (86), Gates (99), Meek (111)

however, that most of these results were based on studies of adults who had already learned to read, and so the cues they used in recognizing words may not have been the ones they used when they first learned to read. Further, the method by which they were taught to read may possibly have affected their method of recognizing less familiar words. Such being the case, a teaching method meanwhile must take into account all the known facts, giving special emphasis to those derived from studies of young children.

At the present time, the alphabet method of teaching beginning reading has disappeared, but phonetic methods are still largely used. The superiority or inferiority of the latter over non-phonetic methods has not yet been finally determined [see Gray (44) and Gates (39)], but they are being used more and more as an adjunct to a whole-word method rather than as an independent method. In this study a whole-word method was employed, but no phonics were taught. The words were presented in sentences for the following reasons: (*a*) It was desired to impress on the minds of the subjects that reading is a meaningful process; that its purpose is to convey ideas. The sentence is the logical means of doing this. (*b*) The sentence, because it "tells a story," is much more interesting to the child, and hence holds his attention better. (*c*) The written sentence is a more natural unit than the isolated word. The child uses oral language to express ideas. Since leaving the stage of babblings behind, the child has progressed from the single word to the sentence containing several words. But, according to Stern (68), these single words are indeed sentences, their meaning

being derived from the total situation, e.g., "Mama" meant "Mama, lift me up," or "Mama, give me my ball," etc. The single written word, however (except the imperatives), does not fulfill the same purpose, but is largely meaningless. The idea requires to be expressed in several words when it is written. (*d*) Words presented in context are more easily recognized than when isolated. Bowden (86) found that her one six-year-old subject at the end of one month's instruction, recognized 83 per cent (of 53 words) in context and 52.8 per cent out of context; and, at the end of two months, she recognized 88.5 per cent (of 85 words) in context, and 62.5 per cent out of context. Sholty (67) found that her three subjects recognized, respectively, 87, 90, and 82 per cent of the words when they were presented in context, and only 61, 50, and 63 per cent, respectively, when they were presented out of context.

Introduction of Printed Symbols by Means of "Massed Forms"

After the decision to use a word-sentence method, the next problem was to find the best means of introducing it. A normal six-year-old child starts to school with the expectation of learning to read and with a certain interest in word symbols already aroused. Not so a normal four-year-old. His interest in these has yet to be developed. That was part of our problem. The survey of literature on the perception of words revealed a general belief that words are seen as wholes. If children do see words as wholes, just what do they see? It may be that they see the geometric shape of the word. Several of the above writers have spoken of

the geometric form of the word, but no experimental work has as yet been done to find out if this is actually a factor in recognition.[3a]

It does not seem to be altogether unreasonable, however, to suppose that such is the case, and a further fact that adds credence to this possibility is that children can recognize a word presented upside down as readily as right side up. This was noted by Bowden (86) and again in the present study. That this is possible surely argues that a word must have for the child a very definite shape. Accepting this as a possibility, the question immediately arose as to whether children with a mental age of four years could discriminate between geometric forms. It is known that they can. According to the Stanford Binet Test a normal four-year-old child should be able to discriminate correctly at least seven out of ten geometric figures used in this test. Here, then, was something that a normal four-year-old child could do, and it was at once decided to make this the starting point of our method. All but one of the subjects selected for this experiment had successfully passed this test. Accordingly, a series of what were called "Massed Forms" was evolved, which served to bridge the gap between no knowledge of printed word symbols and knowledge and interest in them. This might very readily be called a genetic approach to reading.

[3a]In October, 1926, Gates (37) published an article which presents data on the perception of geometric forms. The geometric forms he used, however, were outline forms, and bore no relation to the shape of words.

Description of the "Massed Forms" and the Method of Using Them

The series of massed forms consisted of ten charts of white bristol-board. On the first chart (Figure 1) six simple geometric forms were drawn and washed over with black India ink. A duplicate set of forms was cut out of black cardboard. These will be known as the "sample" forms. On the second chart (Figure 2) four oblongs of different lengths were drawn and massed solid in black as before. The oblong was chosen as the form most closely resembling Messmer's linear words, that is, words containing no projecting letters, such as *can, mouse,* etc. Again the duplicate set was cut out of black cardboard. On the third chart (Figure 3) six oblongs were drawn, but besides being of different length they had projections on the upper and lower sides. This was an attempt to approach the appearance of words with projecting letters, such as *dance, cake, may,* etc. Again the sample forms were cut out of black cardboard. On the fourth chart (Figure 4) seven massed forms were drawn. These were exact representations of actual words, and were made in the following way: With the aid of a printing set, Type No. 1, a word was printed. By means of a ruler and pencil a line was drawn along the top and bottom of the word, but the contour of the front of the first letter and the far side of the last letter, and of projecting letters, was roughly followed. This outline was then transferred to white bristol-board and massed solidly in black as had been done with the previous forms. In this way the rough geometric form of a

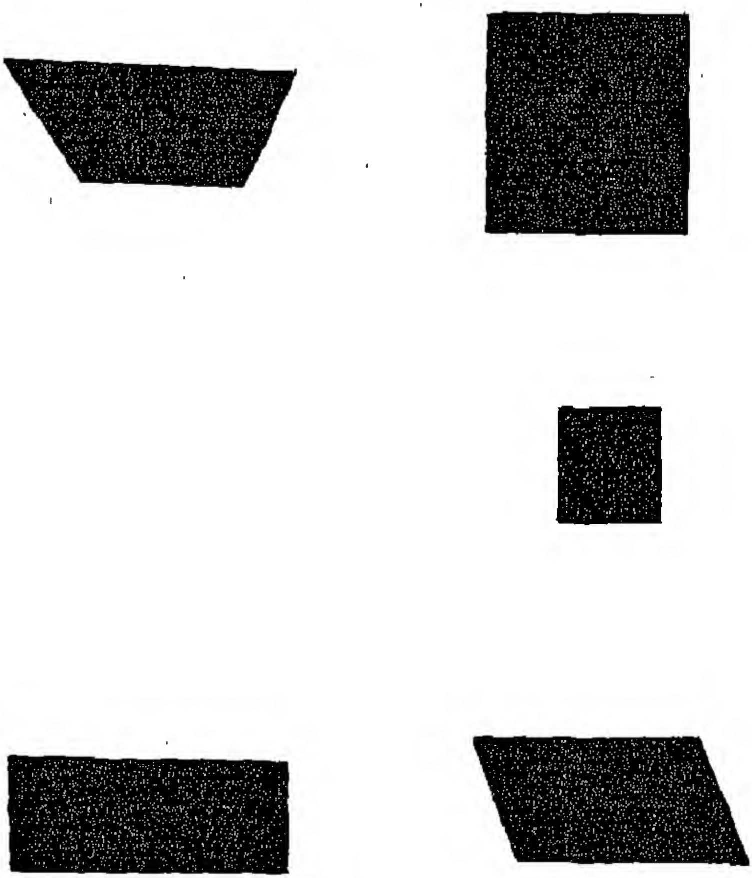


CHART 1
MASSED FORMS
(Reduced one-third)

word was obtained. Each sample form of this set and of the remaining sets was drawn in the middle of an oblong card of white bristol-board. Accordingly, they differed from the sample forms of the previous sets in that they were surrounded by a white margin. On the fifth and sixth charts were massed forms similar to those on Chart 4, but printed by means of a much larger type in order to emphasize the projections. There were only three massed forms on Chart 6, the first two ending in the same letter, and the last two beginning with the same letter. The purpose here was to force the subject who depended solely on using as his cue either the beginning or the end of the massed form to use another cue. The seventh chart contained the names of the subjects in massed form, while the eighth

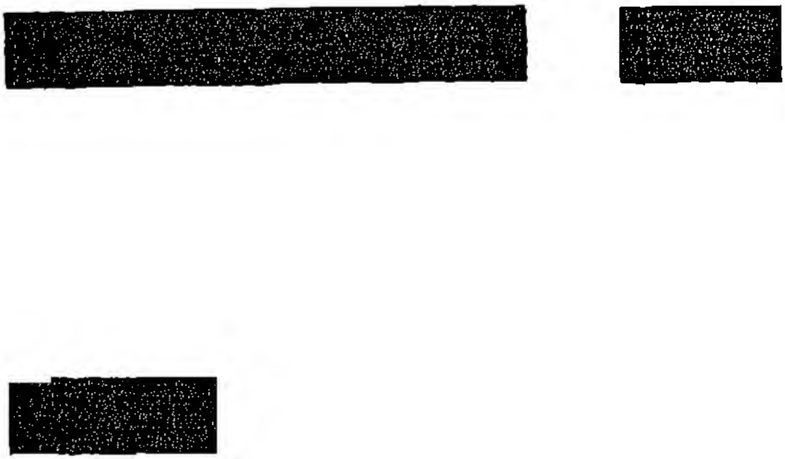


CHART 2

MASSED FORMS

(Reduced one-third)

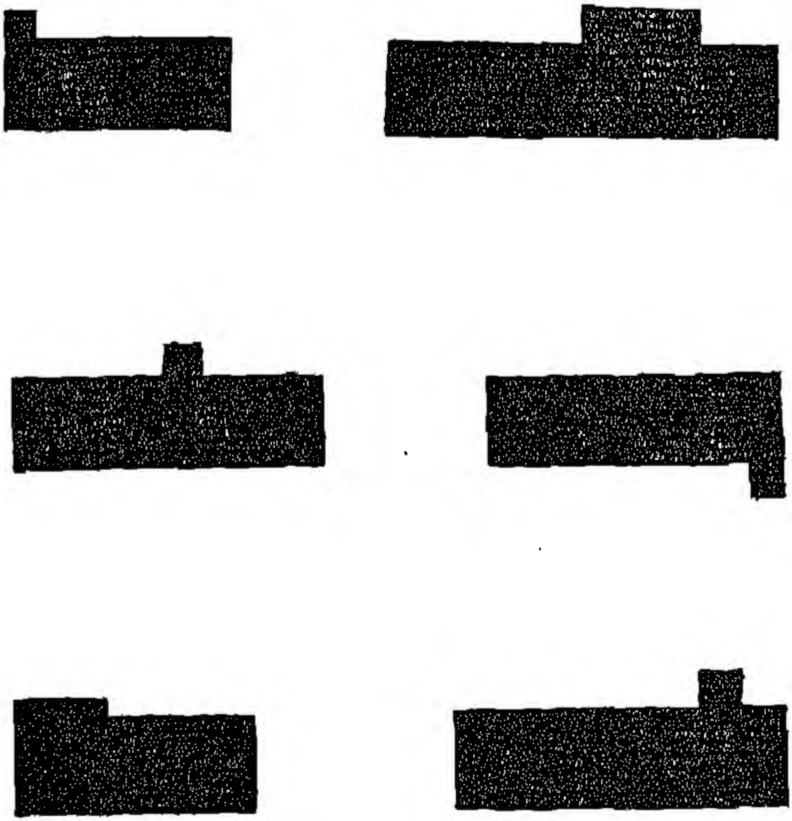


CHART 3
MASSED FORMS
(Reduced one-third)

chart abandoned massed form for real words. The type used in these last two charts was the same as that used on chart four. The words used for the transition series were the children's own names, but they were printed in a different position on the chart to that occupied by their massed form. On Chart 9 were the words *door, chair, table* in massed form, and on Chart 10 these same words in real print. These last two charts were used in connection with the action sentences to be described later.

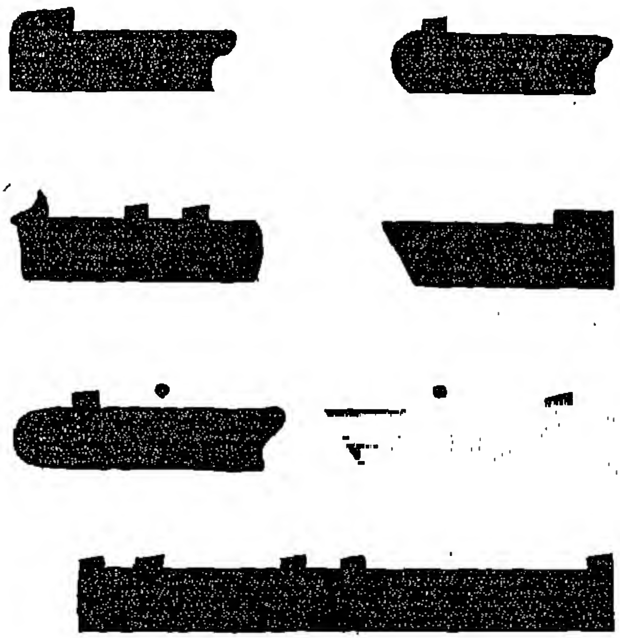


CHART 4
MASSED FORMS
(Reduced one-half)

The method of presentation was essentially the same for each set in the series. In the case of the first three sets, the chart was placed on the table before the subjects. One of the sample forms was placed below the chart on a white background. A subject was asked to find a form on the chart exactly like the sample form. Since these sample forms were cut out exactly, he could place them on the top of the forms on the chart and find out for himself whether or not his selection was correct. Charts 4 to 9 were pinned on the wall, but the procedure was exactly the same as that just described. The sample forms accompanying these forms, however, could not be placed on the top of the chart forms since they were surrounded by a white margin. This made the task of discriminating somewhat harder. In the case of error, the attention of the subjects was drawn to the shape of the beginning and of the end of the form, and to any projecting part, the experimenter saying, "See how this goes," or, "See how this sticks up." In this way definite training was given the subjects in the use of cues that might aid them in recognizing the different forms. Chart 8 was presented immediately after each subject had made two correct discriminations with Chart 7. Nothing was said to cause the subjects to think that these symbols were different in any way from those in the former series. Exactly the same procedure was used as with the massed forms. Each subject again made two correct discriminations. For his second discrimination the experimenter saw to it that each child found his own name. When he had done this he was told that he could hold his card until the others had their turn. The experimenter now said,

"I have a big surprise for you. Do you know what that card, which you are holding, says? It says your name. Now, Kate, look at your card. What does it say?" "Kate." "Now look at these names (pointing to the chart) and find one just like it that says Kate." And so on with each of the other subjects.

Purpose and Value of the Massed Forms

The purpose and value of this use of massed form may be described as follows:

1. It was a simple, logical, gradual, and interesting means of introducing printed word symbols to very young children whose interest in such had not yet been aroused.
2. It gave definite training in discrimination. The habits and skill gained in comparing two massed forms seemed to carry over readily to real words, and proved of considerable value in the reading situation.
3. It gave definite training in looking for and making use of cues. The chief cues used were those of length, the shape of the beginning and of the end of the massed forms, and any pecularities in the middle of the forms. This method was carried over and applied to real words, and the skill and habits established by practice with massed forms seemed to prove of considerable value when real words were involved.
4. It developed a certain amount of independence.

How important a rôle the geometric shape of the word plays in its recognition, if any at all, and how closely ability to discriminate these massed forms is related to ability to discriminate real words cannot be determined here without further experimentation.

Nevertheless, this novel approach to reading was greatly enjoyed by all of the subjects. They entered into these games with great eagerness and evident enjoyment, and the transition from massed form to real words was made so easily that they were discriminating real words without being aware that they were words. Further, as already noted, the habits of discrimination and skill in the use of cues gained from practice with the massed forms seemed to be of definite value in the real reading situation.

This geometric approach would seem also to fulfill the requirements of a Gestaltian psychology. The first series of massed forms might very readily be looked upon as an undifferentiated mass on a white background. Gradually and systematically, this undifferentiated mass assumes a more definite form, a form more and more closely resembling the shape of true words, until, at last, it evolves into the definite letters of a definite word.

As soon as the printed word was introduced, it was used in a variety of situations, such as the morning greeting, the bulletin board, and so on. These provided further means of arousing interest in printed symbols and for demonstrating that reading is a meaningful activity. These various schemes, which developed concurrently and which were abandoned as soon as they had completed their usefulness, will now be described in turn.

The Morning Greeting

Each morning, in response to the experimenter's "Good morning, ——," each subject replied somewhat

formally, "Good morning, Miss Davidson." On the first day of the second week the experimenter used a flash card with the words "Good morning" on it, but continued to name the subjects verbally. On the day following the introduction of the subject's own printed names in the manner described above, the whole greeting was given by means of cards. The subjects had to watch and respond only when they saw their own names. This method was continued until the day of the twelfth reading lesson, when it was abandoned as having fulfilled its purpose.

The Bulletin Board

Each morning, a "story" consisting of a single sentence was printed in white chalk on the section of the blackboard set aside as a bulletin board. The experimenter read it to the subjects and they repeated it after her. Occasionally, one subject was called upon to repeat it all by himself. No attempt was made to make this a reading lesson. Its purpose was merely to arouse interest in printed symbols, and to pave the way for the introduction of known words at the appropriate time. At first, such interesting announcements as "We shall play ball today," were used, the game being duly played later in the day. On the day following the introduction of the printed names already described, and on succeeding days, the story was about one of themselves, e.g., "Elsa crayons nicely," "Paul has on a new suit," etc., thus appealing to the little egoists' interest in themselves. The subjects had to find out for themselves whom the story was about, then the experimenter read the rest of the story to them. As soon as new

words were learned they were introduced into the bulletin story, until, eventually, the subjects could read the whole story with a little help. It never exceeded a single sentence. The bulletin board was continued until April 1, 1926, for groups A and D, and until January 19, 1927, for Group B. It was used somewhat irregularly towards the last.

Applied Seat-Work

Two carefully planned pieces of seat-work,[4] involving the discrimination of their own names and giving practice in the use of cues, were devised for the subjects. The first of these was given on the day following the introduction of the printed names by means of massed forms.

As previously stated, the chief purpose in the early days was to arouse an interest in, and a love for, reading, and to impress on the subjects that reading was a meaningful process. At the same time it was very desirable to develop a certain amount of independence in seat-work in order to free the experimenter for the main part of the experiment. As a means of achieving these ends, it was decided to train the children to do seat-work according to written directions. As this would be largely confined to drawing and coloring, a knowledge of the names of the colors would be important. After the games of sorting names just referred to, a series of games of sorting colors and color names was planned. These served to familiarize the subjects with the color names and gave still further practice in

---

[4]A description of the name- and color-sorting games will be found in the appendix.

matching words. At the close of these games, a larger color chart was pinned up permanently on the wall, low enough to permit the subjects to match the color names with those on their direction sheet. The subjects were always expected to find out for themselves the colors they were to use, and thus a certain amount of independence was developed. The names were printed in both large and small type.

Once the main reading lessons ("the reading lessons proper" to be described later) were begun, seat-work according to written directions was introduced. This generally involved the drawing of an animal by means of stencils, the coloring of it, and, later, the cutting out and pasting of it on colored paper. The reading of the directions was included as part of the regular reading lesson.

Group Games

After the series of lessons constituting the main reading lessons was begun, four or five minutes were usually devoted each day to some form of group game. This was continued until the last month, when the frequently recurring "Vocabulary Tests," which were now rather lengthy, took so much time that it was found advisable to drop the group games. The game most frequently played was the Group Recognition Game. For this game the children were seated in a row. The experimenter exposed the word cards one at a time, and the subject who was first to name the word correctly was given the card. The game was to see who would get the most cards.

Cancellation Games.[5] Occasionally, for a change, each subject was provided with a slip of paper on which a list of words was printed in small type. The instructions were, "Find every word that says 'can' and draw a line around it." These exercises proved very enjoyable and were done with great rapidity.

Action Sentences. Beginning with the second week, an approach was made towards the use of action sentences. Charts 9 and 10 of the massed form series served to introduce the words *door, chair,* and *table.* Sentences, such as "Run to the door," "Hop to the table," were then read and executed by the subjects. By this time, the picture-word vocabulary, which was being developed by means of the picture-language lessons to be described later, contained several words. It consisted of pictures of animals with their names underneath, pasted on the blackboards within easy visual range of the subjects. These words were now made use of in the action sentences. The subjects were asked to "Run to the cow," "Hop to the horse," etc. The action sentences, as a separate lesson from the main reading lesson, were abandoned on February 19, for Groups A and D, and on December 3, for Group B. By this time, the main reading experiment had been established, and it was decided that a daily 10-minute reading lesson for each subject, with perhaps a few minutes group work, would be quite sufficient for these young subjects.

Picture Study Leading to the Main Reading Lesson.

---

[5]A description of the cancellation games will be found in the appendix.

On the third and succeeding mornings, an informal talk was held about a picture. On the seventh morning, a picture of a dog was pinned up. Underneath, on a strip of white bristol-board, was printed in large letters, "This is a dog." After the usual informal talk, attention was drawn to the sentence. The experimenter read it and had the subjects repeat it after her. She "framed" the word *dog* with her hands and said, "That says 'dog.'" Nothing further was said about it. A similar procedure was followed the next day with a picture of a cow and the sentence, "This is a cow." The attention of the subjects was then drawn to the side blackboard on which had been pasted the picture of the dog with the single word *dog* underneath it. On each of the next three mornings, a horse, a rabbit, and a girl, formed the subject of the lesson, and in each case the stereotyped sentence, "This is a ——" was used. These pictures and words formed the nucleus of the picture-word vocabulary.

These picture-language lessons were paving the way for the first real reading lesson. They served several purposes. (*a*) They made use of the children's natural interest in pictures. (*b*) They served as a convenient starting point for conversation. (*c*) They broke down the children's reserve and encouraged them to contribute information and personal experience. (*d*) They gave practice in the use of language. (*e*) They helped to increase the children's vocabulary and to add to their background of common experience. (*f*) The children learned to listen while one child was talking. (*g*) The introduction of the printed sentence provided another occasion for getting acquainted with printed

symbols. (*h*) The same sentence form was used to simplify the first reading lesson.

The development of a picture-word vocabulary was also for a definite purpose. It furnished the subjects with another opportunity to become familiar with printed symbols and to apply their newly acquired ability of matching words. It developed a very desirable independence, for no subject was ever told a word that appeared in the picture-word vocabulary. He was expected to find it out for himself by putting into practice his ability to note points of resemblance and of difference. This proved to be a very delightful and valuable game.

The Reading Lesson Proper

These various avenues of approach that have just been described were both of an introductory and of an exploratory nature. They served to introduce printed symbols to the subjects and to arouse an interest in them, thus paving the way for the main reading lessons. The first of these was not given until the *thirteenth day of meeting*. They also enabled the experimenter to test out the capacities and limitations of these young subjects, and thus permitted a more careful planning of the details of the reading method proper.

In planning these avenues of approach and the reading lessons proper, the experimenter constantly kept in mind the discrepancy in age between these subjects and the average six-year-old beginner. A difference of two years in mental age is a very large difference in these early years. A normal four-year-old child has not left babyhood far behind, his experience of this

world is still very limited, his motor coordination is not very well developed, his attention is very fleeting, and his capacity for voluntary attention is extremely small. A teaching method suitable for a normal six-year-old subject might not prove suitable for a four-year-old one. A process that was easy for a six-year-old subject might prove to be full of difficulty for the four-year-old one. Every effort was made to avoid physical and mental fatigue, strain of the eyes, and boredom. Above all, it was desired to arouse a love of books, and stories, and reading, and to avoid everything that might be inimical to these. For these reasons, the experimenter attempted to foresee what might prove to be a difficulty for a normal four-year-old subject and so analyzed the reading situation into its very simplest elements according to her best judgment. Five clearly defined stages of possible difficulty were noted, and the lessons were planned so as to introduce these stages gradually. This is an extreme simplification of reading method destined to avoid confusion, fatigue, and eye-strain, until proper eye-movements are established and words are comprehended as units. A brief description of these so-called stages of difficulty, together with reasons for so regarding them, will now be given.

Stage 1. (*a*) The words to be used in a sentence were presented separately on cards, word by word. The subject was required to name each word before he placed it in the reading rack. (*b*) The sentence was built up from left to right. (*c*) Only one sentence was built up in the reading rack. (*d*) As soon as it was thus "built up" the subject read the whole sentence.

(*e*) The sentence was then taken down in irregular order.

The reasons for introducing the first sentences in this fashion are as follows: In the first place, the young child does not know what a word is. To present him with a whole sentence at once, or a whole nursery rhyme, as is commonly done nowadays, is assuming that he has certain abilities which it is not known for a certainty he possesses. It assumes that the amount of white separating two words is sufficient to convey to him, without instruction, the idea that these are separate units. This may or may not be the case. It assumes that he has the capacity to attach one word name to each word symbol in the sentence or rhyme. This seems to be quite a large assumption to make in the case of the normal four-year-old. The reading situation is not unlike the counting and tallying situation. According to the Stanford Binet, the majority of normal four-year-old children should be able to count four pennies, pointing to each coin in turn as he counts. Only 5 of these 13 subjects had passed this test. In reading, for example, "Mary had a little lamb," he has to attach a name to each printed symbol as in counting. But it logically appears to be more difficult than counting, for some words contain two syllables. This means, then, that in some cases, one printed symbol stands for a one-syllable word name and sometimes for a two-syllable word name. It was thought that this might prove a source of confusion for these subjects in the very earliest stages unless a little help was given. Accordingly, only one word was presented at a time, and the word name correctly associated with this

printed symbol, whether it were a one- or two-syllable word, before another word was introduced. Each word being on a separate card emphasized the unity of that particular word symbol. Another advantage that this method seemed to offer was the establishment, right from the very start, of the correct left-to-right eye-movements used in reading. A child has to learn to begin at the top left-hand corner and to proceed line by line from left to right. A good teaching method will help him to gain these habits in the simplest possible way. To present him with several lines of meaningless print must not only be somewhat bewildering but also rather fatiguing until these eye-habits are established.

Since Buswell (87) and other investigators have found that the eyes of first-grade children make several pauses with each word, it was felt that the presentation of one word at a time would be an advantage rather than a disadvantage in these very first stages. It would avoid peripheral stimulation which, though a very valuable aid to speedy reading, might prove to be a distracting factor until greater familiarity had been gained with several words. Of course, peripheral stimulation was not entirely eliminated. It had an opportunity to exert its influence when the sentence was read after having been "built up"—a very desirable feature. But there was no enticement to let the eyes wander all over the page, for only one sentence was presented until the eye-movements had become more stabilized and increase of vocabulary had given a greater degree of self-confidence and of self-control.

Stage 2. A sentence was built up in the rack as before, each word being named by the subject before it

was placed in the rack. This time, however, it was not read until a second sentence had been built up in a similar manner underneath it. Then both sentences were read, one after the other.

Now that considerable practice had been gained in reading single sentences, this was only slightly harder than Stage 1. It entailed ability to retain words for a longer period than before, and ability to resist the distraction of the second line until the first had been read. More important than either, perhaps, was the introduction of the eye-movement from the end of one line to the beginning of the next line.

Stage 3. Two sentences were built up together in the rack as in Stage 2, but with this difference: The subject was not called upon to name each word before he placed it in the rack as before, but he was told just to look at it (read it silently). He then read the two sentences aloud. This was in preparation for Stage 4.

Stage 4. A whole sentence was presented at once without building it up with separate cards. This sentence was read before a second one was added. This assumed that the children had now learned to appreciate a word as a unit; that they could recognize certain words with greater ease; that their eye-movements had become more stabilized. It also permitted the normal influence of peripheral stimulation.

Stage 5. Several sentences were presented at once on a chart. This was the final stage before beginning the use of a primer. It was early noticed that the transition from large type to small type was a real difficulty with these subjects, and definite help was therefore given them to make it easier. This was begun in

connection with the reading of written directions, and so paved the way for the introduction of the primer.

Subject-Matter

For subject-matter the experimenter made use of the fact that children of this age are intensely interested in themselves and in children like themselves. Several real children and the subjects themselves were made the center of the simple reading lessons. The early lessons were "built up" in the reading rack which was pinned up on the wall, and the lessons were timed with a stop-watch.

During the pre-primer period, the 234 sentences given in the appendix were read by all the subjects. It will be remembered that five stages of possible difficulty had been noticed. Sentences 1 to 36 were presented by the method described in Stage 1, sentences 37 to 42 as in Stage 2, sentences 43 to 53 as in Stage 3, sentences 54 to 82 as in Stage 4, and sentences 83 to 234 as in Stage 5. Wherever possible, a picture was made the starting-point of the lesson. A piece of seat-work appropriate to the reading lesson was generally planned for each subject. This frequently, but not always, involved reading. It usually consisted of the drawing, coloring, and cutting out of a picture, or of plasticine work. When written directions were involved, they were first built up in the reading rack and then read in both large and small type as part of the regular reading lesson. In this way definite help was given in making the transition from large to small type. Groups A and D used stencils, but these proved too difficult for Group B, whose motor coordination was not so well

developed, and so the experimenter made the necessary drawings for this group, leaving them to do the coloring and cutting out.

It may be of interest to describe in detail the method used in an early lesson, and to point out its principal features. Lesson 5 has been selected for this purpose.

LESSON 5 (Sentences 12 to 18)

a) Review of words.

b) Reading.

1. The words *Janet, can,* and *run* were taught.

2. Sentence 12 was "built up" in the following way. *E*[a] held up a card on which was printed *Janet*. *S* named it, and placed it in the reading rack. Similarly, he named *can* and *run* and placed them in the rack in proper order. *S* now read the whole sentence. *E* now said, "Give me the card that says 'can,' 'run,' 'Janet.'" *S* selected the proper card and handed it to her. The other sentences were built up, read, and taken down in a similar fashion. The following points are to be noted:

1. At the beginning of each lesson a brief review of old words was given and new words were drilled upon. Attention was drawn to the beginning and end of a new word and to any distinguishing feature of it which could be used as a cue until it was throughly assimilated. (Later on, when the vocabulary had become somewhat large, the lesson was divided into two five-minute periods, separated by an interval of time. The

[a]The capital letter *E* stands for the experimenter, and *S* for the subject.

first period was spent in review and drill of words and the second in reading stories.)

2. Drill on the words out of context helped to unite name and printed symbol.

3. Each word was recognized and named before being put into the reading rack.

4. The sentence was built up in the correct left-to-right order. This made the sentence easy to read, and afforded a chance for memory to play a part.

5. Building up in order helped to establish the left-to-right eye-movements used in reading.

6. The "taking down" in any order put the emphasis on the correct recognition of each word and prevented entire reliance on position in the rack. That "position" was a factor was frequently noted.

7. The child's natural delight in handling material was made use of by having him place the cards in the rack himself.

8. In sentence 13, and in others whenever possible, his activity was made use of in a more vigorous way, viz., in running.

9. The context was well within his comprehension and experience and held his interest. The introduction of his own name always proved to be very pleasing to the subject.

10. As soon as a new word was introduced, it was used in a variety of sentences and situations. This was done both to increase its meaning-value and to bring about its recognition in any context. It will be noted that, in sentence 13, the word *can* printed with a small *c* was used at the beginning of a sentence, contrary to customary usage. To these subjects *Can* with a capital

letter was a different word from *can* with a small letter. Three new words had already been introduced and it was felt that this was enough for one lesson. On the other hand, it was desirable to use the word in different situations for the reasons stated above, hence the adoption of this procedure. Words which were to be used chiefly at the beginning of a sentence were always printed with a capital. Later on, both forms were introduced.

11. Use was made of the picture-word vocabulary pasted on the side-blackboard. *E* never told *S* a word that was pictured.

Summary of the Chief Features of the Reading Method

1. Use of a word-sentence method.
2. Introduction of the printed word symbol by means of a graded series of massed forms.
3. Definite training in the use of cues which might be helpful in discriminating words.
4. Development of independence through the use of a picture-word vocabulary.
5. Use of the child's love of motor activity.
6. Avoidance of fatigue through a special method of introducing the sentence.
7. Development of proper eye-movements.
8. Gradual introduction of difficulties.
9. Gradual introduction of small type.
10. Some slight use of group competition in games.
11. Use of a syllabic method of teaching derivatives of known words and of compound words.
12. No use of phonics.

IV

ACCOMPLISHMENT OF THE SUBJECTS AS MEASURED BY VOCABULARY AND READING TESTS[7]

A. Tests Given during the Course of the Experiment

1. *Results of the Vocabulary Tests Including the Final and Retest.* The following vocabulary records were obtained for each subject: (*a*) Five daily vocabulary records for the fifth to the ninth lesson inclusive. (*b*) A vocabulary record taken on the day of the 21st reading lesson, irrespective of the actual number of reading lessons each subject had received. For convenience, this will be called Test 21. (*c*) Complete vocabulary tests given to each subject as he reached his 30th, 40th, 50th, and 60th reading lesson. These will be known as Vocabulary Tests 30, 40, 50, and 60. These tests should be very comparable since each subject had received exactly the same number of reading lessons. Only one subject, Sally of Group B, was given the 60th test. On the next day she was given a regular reading lesson, and on the following day the final tests. Two subjects, George and Bell of D, reached their 60th lesson on the day preceding the first day of the final tests. As lengthy vocabulary tests were always spread over two days, it was decided to let the 60th lesson count as the final test in their cases. In a similar way, for Edgar of D the 50th test is counted as

[7]A description of all the tests used in this experiment and the method of administering them will be found in the appendix.

the final test. Dora of A was absent between Test 40 and the final tests.

Tables 4, 5, and 6 present the results of these tests. As it is desirable to discuss the results of all the vocabulary tests at the same time, the results of the final vocabulary test and of the retest have been included in Table 6. In this table are also included the actual number of lessons received by each subject, and the average number of words learned in each lesson as represented by the Final Vocabulary Test score.

From these three tables it is seen that in every test

TABLE 4

FIVE EARLY VOCABULARY RECORDS

	Bright					Average				Dull			
Record	Jean	Paul	Peter	Kate	Sally	Leo	Elsa	Lela	Dora	Geo.	Edgar	Don	Bell
1.	4	4	5	2	5	4	5	ab	2	4	5	3	2
2.	7	ab	ab	2	8	5	4	4	ab	8	5	3	4
3.	9	6	9	2	8	9	7	7	ab	9	8	4	4
4.	8	6	10	3	10	6	4	8	4	10	7	2	5
5.	12	6	10	ab	13	3	8	11	4	11	ab	3	7
Total	40	22	34	9	44	27	28	30	10	42	25	15	22
Av.	8.0	5.5	8.5	2.3	8.8	5.4	5.6	7.5	3.3	8.4	6.3	3.0	4.4
Grand average			6.6					5.5			5.5		

TABLE 5

VOCABULARY TEST 21

Bright				Average				Dull			
Subj.	Score	Lessons	Words per lesson	Subj.	Score	Lessons	Words per lesson	Subj.	Score	Lessons	Words per lesson
Jean	28	18	1.60	Leo	19	19	1.00	Geo.	19	20	.95
Paul	12	17	.70	Elsa	22	14	1.60	Edgar	22	17	1.30
Peter	18	17	1.10	Lela	28	18	1.60	Don	13	19	.70
Kate	6	18	.30	Dora	12	16	.75	Bell	16	18	.90
Sally	39	20	1.95								
Av.	20.6	18.0	1.13		20.25	16.75	1.24		17.5	18.5	.96

TABLE 6

VOCABULARY TESTS 30, 40, 50, 60, FINAL TEST AND RETEST

	Bright					Average			
Test	Jean	Paul	Peter	Kate	Sally	Leo	Elsa	Lela	Dora
30	47	29	42	33	64	25	47	38	20
40	82	43	57	45	120	30	56	44	21
50	156	41	81	52	185	29	70	59	—
60	—	—	—	—	250	—	—	—	—
Final	191	56	81	50	269	23	98	75	25
Retest	120	18	18	21	348	3	52	16	9
No. of lessons	55	51	52	51	61	59	55	55	40
Av. words per lesson (Final Test)	3.5	1.1	1.6	1.0	4.4	0.4	1.8	1.4	0.6

	Dull				Mean group scores		
Test	Geo.	Edgar	Don	Bell	B	A	D
30	30	29	14	20	43.0	32.5	23.3
40	51	39	16	27	69.4	37.8	33.3
50	60	40	17	34	103.0	52.7	37.8
60	—	—	—	—	—	—	—
Final	66	40	20	34	129.4	55.3	40.0
Retest	8	—	11	8	105.0+	20.0	9.0
No. of lessons	60	50	53	60	54	52	56
Av. words per lesson (Final Test)	1.1	0.8	0.4	0.6	2.4	1.06	0.7

Group B secured the highest scores, Group A the next highest scores, and Group D the lowest scores. In Test 21, when the number of lessons actually received was taken into account, it is seen that Group A was slightly superior to Group B. From Test 30 on, however, Group B became increasingly superior to the other two groups, as shown by the mean group scores given in Table 6, until, in the final test, it excelled Group A by 134% and Group D by 221%. Group A, on the other hand, maintained a superiority over Group

D in all but the retest of Table 6 of slightly under 40%.

The presence of large individual differences is revealed by these tables. When attention is confined to the Final Vocabulary Test, it is seen that scores range from 20 to 269. The two most outstanding scores are those of 269, gained by Sally[8] at the age of three years and four months, and of 191, gained by Jean at the age of three years and eight months. These scores represent a rate of learning of 4.4 and 3.5 words per lesson, respectively. The third highest score, one of 98, was gained by Elsa of Group A. This is a very satisfactory score, although completely overshadowed by the two scores just mentioned. This subject was ill throughout the whole of the last week of instruction, and she was tested in her home although she had not yet fully recovered from her illness. She was extremely restless, and it is felt that she did not do herself justice. The poorest score was obtained by Donald, of Group D, who had shown little interest in reading throughout the experiment. The next poorest score was obtained by Leo of Group A. A singularly inert child at all times, Leo was at his worst during the testing period. The wave of hot weather at this time may have contributed to his listlessness. The third poorest score was obtained by Dora, of Group A. She had dropped out of the experiment for a time, apparently discouraged at her lack of success as compared with others in her group, but, later on, she voluntarily asked to resume her

[8]The attention of the reader is directed to the brief sketches of the subjects and of their attitude to reading which will be found in the appendix.

lessons, and was beginning to show evidence of returning interest and of fair ability when the experiment closed. The scores of Edgar and Bell, both of Group D, were disappointing in view of the ability and interest they generally displayed. Illness in the home towards the end of the experiment caused Edgar to be somewhat irregular in attendance, and no doubt this affected his score. It is of interest to note that Paul of B, who gained a score of 41 in Test 30, succeeded in gaining a score of 56 two days later. This is an increase of 15 words or of 36%. This seems to indicate that moods and health influence the effort expended by preschool children very easily.

At this point it may be well to call attention to the results of the retests. It has already been pointed out that an unequal length of time elapsed between the final tests and the retests. Groups A and D were retested four months after the close of instruction, while Group B was retested after an interval of only nine weeks. As a consequence, the results are not strictly comparable. They serve to show, however, the rapidity of forgetting.

Only one subject, Sally of B, improved her score. She recognized 348 words out of context, but this score by no means did her justice. She had read a great deal since the close of the experiment and the writer had no means of testing her on her newly acquired vocabulary. Her true score would doubtless have been considerably in excess of this. According to an information blank filled out by her parents, she had received no instruction in reading in the interval. She read a great deal by herself, however, and when she came to

a word she did not know, she would appeal for help. Her parents would then pronounce the word for her. This was the only help she received. In this period she had read the following: The Elson, Beacon, and Bolenius *Primers,* the Bolenius *First Reader,* several variations of *Mother Goose, Gordon, Bow Wow and Mew, Five Little Friends, Goody Two Shoes, Child's Garden of Verses,* and parts of several other books—no mean record for a child to achieve before the age of three years and seven months.

Jean succeeded in retaining a considerable portion of the words she had formerly known. She had read a little in her sister's primer in the interval, but had received no instruction. She had possibly been handicapped by upset home conditions, for there was illness in her home throughout the entire experiment, in addition to her own illnesses.

According to information blanks filled out by all the parents, only two subjects had received regular instruction since the close of the experiment. These were Elsa and Lela, both of Group A. Elsa, who had retained about 50 per cent of her vocabulary, had been given about two hours instruction per week for a period of six weeks. She read one page of her primer each time. Lela received from one-half to one hour's instruction every few days for a period of eight or nine weeks. She read her primer and the children's pages of magazines. She retained only 20 per cent of her vocabulary. Leo had forgotten all but three words, and Donald, who had entered first grade after the summer vacation, recognized only 11 words despite school instruction.

2. *Results of the Unstandardized Reading Tests.* In addition to the vocabulary tests described, each subject was given an unstandardized reading test on the day of his 40th and 50th reading lesson. The results of these reading tests[9] are presented in Table 7.

As in the vocabulary tests, Group B gained the highest and Group D the lowest scores in each of the reading tests. The superiority of Group B was greatest in the hardest passages of Test 50 (II and III), but it is not quite so striking as in the vocabulary tests. This table, too, reveals the presence of large individual differences. Jean of Group B read three of the four tests without error, and Sally read two without error. These two subjects read the stories with great rapidity and confidence. The majority of the subjects, however, were still in the word-naming stage. Elsa of Group A had been on the threshold of fluent reading when her progress was interrupted by illness, while Peter of B, Lela of A, and George of D were close to the threshold. By fluent reading is meant ability to read an unknown story of familiar material independently, rapidly, and with few errors. It does not refer to the rereading of old stories.

B. Final Tests and Retests

The results of the vocabulary tests given at the close of the experiment and again at the time of the retests have just been described. The results of the standard-

[9]Time records were kept for all reading tests, but, since the amount of time allowed for the recognition of a single word was not kept constant, they were found to be rather unsatisfactory for all but the most fluent readers, and so they have not been included in the table of results. This applies, also, to the final reading tests.

TABLE 7

UNSTANDARDIZED READING TESTS 40 AND 50

(Score = number of words read correctly)

Total possible score	Test 40 17	Test 50-I 16	Test 50-II 18	Test 50-III 18	Total Test 50 52
Bright					
Jean	17	16	18	16	50
Paul	15	6	7	6	19
Peter	16	13	13	11	37
Kate	13	8	9	8	25
Sally	16	16	18	17	51
Av.	15.4	11.8	13.0	11.6	36.4
Average					
Leo	13	9	6	7	22
Elsa	16	12	16	13	41
Lela	15	13	11	9	33
Dora	12	—	—	—	—
Av.	14.0	11.3	11.0	9.7	32.0
Dull					
George	15	12	12	9	33
Edgar	15	13	9	9	31
Donald	6	6	4	4	14
Bell	12	10	10	7	27
Av.	12.0	10.2	8.7	7.2	26.2

ized and unstandardized reading tests given at the close of the experiment are presented in Table 8. The scores in all but the Detroit Word Recognition Test and the Pressey First Grade Reading Test represent the number of words correctly read. The Detroit and Pressey Tests were scored strictly according to the directions given in their respective manuals. A second column of scores has been given for the Detroit test in this table. It is marked "Score at the end of unlimited time," and is included for the following reason. Primary-grade children are given four minutes in which to do as much of this test as they can. The test consists in drawing lines from words and phrases to

the pictures which they represent, and it was thought that the time limit might handicap the subjects of this experiment since their motor coordination was not so well developed as that of normal first-grade children. Accordingly, they were given as much time as they wanted, but the writer made note of the point they had reached at the end of four minutes in order to have data strictly comparable with the norms for this test. As it turned out, only one score was significantly, and two scores very slightly, changed in the final test, and one score was significantly changed in the retest.

From Table 8 it is seen that Group B is superior to

TABLE 8
SCORES IN FINAL READING TESTS

	Improvised Tests		Wag and Puff Test		Merrill Tests			
	I	II	Reading test	Comprehension	I	II	III	Total
Total possible score	14	23	46	4	24	26	33	83
Bright								
Jean	14	22	46	4	21	24	30	75
Paul	11	6	21	3.5	6	6	5	17
Peter	12	15	31	4	10	12	13	35
Kate	12	10	24	3	9	7	11	27
Sally	14	22	45	4	23	25	31	79
Av.	12.6	15.0	33.4	3.7	13.8	14.8	18.0	46.6
Average								
Leo	7	6	17	4	1	7	4	12
Elsa	13	16	34	3	13	9	19	41
Lela	9	13	30	3.5	10	12	12	34
Dora	8	8	12	1	4	6	9	19
Av.	9.3	10.8	23.3	2.9	7.0	8.5	11.0	26.5
Dull								
George	12	9	24	3	7	8	15	30
Edgar	11	11	22	1	12	8	13	33
Donald	6	4	18	2	2	7	3	12
Bell	9	11	22	1	6	7	10	23
Av.	9.5	8.8	21.5	1.8	6.8	7.5	10.3	24.5

TABLE 8 (*continued*)

	Detroit Word Recognition Test		Pressey First Grade Test		
	Score at the end of 4 minutes	Unlimited time	Word	Sentence	Total
Bright					
Jean	15	22	20	14	34
Paul	2	3	6	3	9
Peter	5	5	11	5	16
Kate	1	1	12	2	14
Sally	10	10	21	15	36
Av.	6.6	8.2	14.0	7.8	21.8
Average					
Leo	1	1	7	2	9
Elsa	6	7	12	5	17
Lela	4	4	7	3	10
Dora	2	2	4	1	5
Av.	3.3	3.5	7.5	2.8	10.3
Dull					
George	0	0	10	2	12
Edgar	4	4	11	3	14
Donald	1	1	5	1	6
Bell	1	1	4	3	7
Av.	1.5	1.5	7.5	2.3	9.8

the other two groups in every reading test. Group A is superior to Group D in every test but Form I of the Improvised Test, when Group D is very slightly superior. Not only does Group B read more accurately than the other groups, but it displays greater comprehension of what it reads as measured by the questions asked in connection with the Wag and Puff Test. It is interesting to compare the achievement of the two most successful subjects in the Wag and Puff Test with that of first-grade children. According to the manual (48), first-grade children who have finished reading the primer and have had 17 weeks of instruction (85 daily lessons) are expected to read this selection in 30 seconds with not more than three errors. Jean of B

read it in 39 seconds with no error, and Sally of B read it in 32 seconds with one error. It will be very readily granted, then, that these two three-year-old subjects were reading quite as well as first-grade children. The very slightly longer time taken by them is more then compensated for by their greater accuracy. And they have reached this point in less time than is usually taken by first-grade children. For, whereas first-grade children are expected to attain this point only after 85 daily lessons, Sally had had only 61 lessons and Jean 55. It is of interest, also, to compare the amount of time spent in instruction for these two groups. Each subject of this group received a ten-minute lesson daily, with an additional few minutes of group work, not exceeding five minutes for these two subjects. Accordingly, it may be said that these two subjects received at most 15 minutes of instruction daily, whereas, according to the manual, first grade children received 90 minutes of instruction daily.

Since norms are available for both the Detroit and the Pressey tests, it is of interest to compare them with the scores obtained by these subjects. The norms for the Detroit Word Recognition Test are given below.

	Beginning of term			End of term		
Grade	Q_1	Mdn.	Q_3	Q_1	Mdn.	Q_3
1B	1	3	4	7	12	19
1A	6	10	14	14	20	27
2B	14	18	25	21	28	34

It is at once seen that the score of 6.6 obtained by Group B is practically the same as the lower quartile norm for the end of Grade 1B. The other group scores are too low for comparison, but several of the individ-

21.8 in the Pressey Test. This is almost five points above the norm for the end of Grade 1B. Its superiority was gained in the Word Test, being just slightly below the norm in the Sentence Test. According to this test, then, Group B was reading as well as normal first-grade children who have finished Grade 1B. Groups A and D gained scores slightly better than half the norm. They approached the norm for the Word Test very closely, but were considerably inferior in the Sentence Test. The arrangement of the latter in long horizontal lines prevented the individual sentences from standing out clearly, and proved very confusing and discouraging to those subjects who had not yet reached a stage of independence which enabled them to attack such a mass of unfamiliar print successfully. The greater success with the Word Test was no doubt partly due to experience gained in the "cancellation" exercise mentioned.

Individual scores ranged from 5 to 36. This last score was gained by Sally of B. It is five points above the norm for the end of Grade 1A. She, unlike the majority of the others, was equally successful with the Word and Sentence Tests. In fact, one might well say she was superior in the latter as she made a perfect score in it, and so her ability in this respect was inadequately measured. Jean's score of 34 is three points above the norm for the end of Grade 1A. She lost one point in the Sentence Test and five in the Word Test. Both children had attained an enviable degree of independence, and were reading, according to this test, as well as normal children in the first weeks of second grade. The third highest score, 17, was gained by Elsa

of A. This is the exact norm for Grade 1B. As previously stated, she and two or three others were hovering on the brink of fluent reading. Peter of B was only one point, and Kate of B and Edgar of D were just three points below the norm for Grade 1B. All of these subjects gained a score greater than the norm for the Word Test, but were considerably inferior in the Sentence Test.

It may be that accurate word recognition was overemphasized in the present method, but there is ample evidence that context received its due share of attention. The subjects clearly showed appreciation of reading as a meaningful and interesting process. They rarely read the same story twice, the intention being to have them meet the same words in as many different situations as possible. This caused the words to gain a very desirable independence of context. Too little emphasis on accurate word recognition, with too great reliance on context or on memory, is apt to lead to the formation of wrong associations which will have to be corrected later on. That our subjects, who have successfully passed what has been referred to as the word-naming stage, have been in no way handicapped by bad habits is clearly seen in the rapidity and great accuracy of reading displayed by them. It will be further demonstrated in the good eye-habits built up as revealed by photographs. Nor has comprehension been sacrificed.

Table 9 presents the results of the retests. As already stated, the results of the retests are unsatisfactory, since the time elapsing between them and the final tests was not the same for all groups. Nevertheless, there are

several interesting points to be noted. Less has apparently been forgotten in the reading tests than in the vocabulary tests, some subjects gaining larger or maintaining the same scores that they had obtained in the final tests. Where losses did take place, they do not appear to be as large as those occurring in the vocabulary tests. Since the latter were given first, it may be that they served to arouse sleeping memories, thus enabling the subjects to be more successful with the reading tests, while the occurrence of the word in context no doubt aided recognition. Five subjects improved their scores in the Detroit Test (score at the end of four minutes),

TABLE 9
RETEST SCORES OF THE READING TESTS

	Improvised Tests		Wag and Puff Test		Merrill Tests			
	I	II	Reading test	Comprehension	I	II	III	Total
Total possible score	14	23	46	4	24	26	33	83
Bright								
Jean	12	19	33	4	15	19	26	60
Paul	5	4	26	3	2	7	5	14
Peter	10	9	11	4	8	9	11	28
Kate	5	7	23	4	4	12	7	23
Sally	14	23	46	4	24	25	33	82
Av.	9.2	12.4	27.8	3.8	10.6	14.4	16.4	41.4
Average								
Leo	3	4	12	1	2	6	5	13
Elsa	10	10	29	3	16	13	15	44
Lela	4	8	17	4	9	8	8	25
Dora	5	7	12	2	1	4	9	14
Av.	5.5	7.3	17.5	2.5	7.0	7.8	9.3	24.0
Dull								
George	6	4	2	0	3	3	5	11
Edgar	—	—	—	—	—	—	—	—
Donald	3	2	15	4	7	10	13	30
Bell	2	3	6	0	4	2	6	12
Av.	3.7	3.0	7.7	1.3	4.7	5.0	8.0	17.7

TABLE 9 (*continued*)

	Detroit Word Recognition Test Score at the end of 4 minutes	Unlimited time	Pressey First Grade Test Word	Sentence	Total
			25	15	40
Bright					
Jean	14	14	15	0	15
Paul	2	2	5	2	7
Peter	1	1	6	6	12
Kate	1	1	5	4	9
Sally	21	28	25	15	40
Av.	7.8	9.2	11.2	5.4	16.6
Average					
Leo	0	0	4	0	4
Elsa	6	6	14	6	20
Lela	6	6	9	3	12
Dora	1	2	1	2	3
Av.	3.3	3.5	7.0	2.8	9.8
Dull					
George	3	3	8	3	11
Edgar	—	—	—	—	—
Donald	3	4	9	3	12
Bell	2	2	5	1	6
Av.	2.7	2.7	7.3	2.3	9.7

and three made no change. The largest gain in this test was made by Sally of B, who doubled her score. It has already been pointed out that her final score of 10 did not do her justice. Her score should, in reality, have been 20, for reasons already given. Accordingly, her gain is not as large as it seems. On the other hand, when she was given unlimited time she raised her score to 28. This is the norm for the end of Grade 2B. At this time Sally was barely three years and seven months old. She gained a perfect score in the Pressey Test. This represents a score nine points above the norm for Grade 1A, no other norms being given. She made a perfect score in all of the other tests except Form 2 of the Merrill Tests, in which she made one error. Jean

of B, who did so well in the final tests, received poorer scores in all the retests. It is felt that she did not do herself justice on this occasion, as she became ill after one day's testing and had to be given all of the other tests on the same day, the day before she returned to her home state. Despite frequent rest periods, she became more and more fatigued as the morning advanced, until the writer finally abandoned the testing before the whole program had been quite completed. Her fatigue was very noticeable in the Pressey Test. She began quite well, but soon resorted to guessing, a thing she had never done before, and ended by failing in all of the sentences. The effects of her illness were also evident in the poor motor coordination she displayed in this test. She found it quite a task to draw the required lines round the words. Anyone comparing the lines drawn on this paper with those drawn at

TABLE 10

RANK POSITIONS OF ALL SUBJECTS IN ALL TESTS

	Group	1 Early rec.	2 Voc. 21	3 Voc. 30	4 Voc. 40	5 Voc. 50	6 Read. T40	7 Read. T50	8 Final Vocab.	9 Improvised	10 Wag and Puff
Sally	B	1	1	1	1	1	3	1	1	1.5	2
Jean	B	4	3	2.5	2	2	1	2	2	1.5	1
Elsa	A	7	3	2.5	4	4	3	3	3	3	3
Peter	B	2	6	4	3	3	3	4	4	4	4
Lela	A	5	3	5	7	6	6.5	5.5	5	6	5
George	D	3	8	7	5	5	6.5	5.5	6	8	6.5
Paul	B	8	10.5	8.5	8	8	6.5	11	7	10	10
Kate	B	13	13	6	6	7	9.5	9	8	6	6.5
Edgar	D	6	5	8.5	9	9	6.5	7	9	6	8.5
Bell	D	10	7	10	10	11	11.5	8	10	9	8.5
Dora	A	11	9	11.5	11	10	11.5	—	11	11	13
Leo	A	9	10.5	11.5	12	—	9.5	10	12	12	12
Donald	D	12	12	13	13	12	13	12	13	13	11

TABLE 10 (*continued*)

	Group	11 Merrill	12 Detroit	13 Pressey	14 Grand Av. (1-13)	15 Average of finals (8-13)
Sally	B	1	2	1	1.3 (1)	1.4 (1)
Jean	B	2	1	2	2.1 (2)	1.6 (2)
Elsa	A	3	3	3	3.5 (3)	3.0 (3)
Peter	B	4	4	4	3.8 (4)	4.0 (4)
Lela	A	5	5.5	8	5.5 (5)	5.8 (5)
George	D	7	13	7	6.8 (6)	7.9 (8)
Paul	B	11	7.5	9.5	8.9 (9)	9.2 (9)
Kate	B	8	10.5	5.5	8.4 (8)	7.4 (7)
Edgar	D	6	5.5	5.5	7.1 (7)	6.8 (6)
Bell	D	9	10.5	11	9.7 (10)	9.7 (10)
Dora	A	10	7.5	13	10.8 (12)	10.9 (11)
Leo	A	12.5	10.5	9.5	10.0 (11)	11.4 (12)
Donald	D	10.5	12.5	12	12.3 (13)	12.0 (13)

the time of the final tests would have difficulty in believing that they had been done by the same child.

Rank Positions of All Subjects in All Tests

Table 10 is of more than passing interest. In it the rank of every subject in every test is given. The names are arranged according to the ranking of the subjects in the Final Vocabulary Test.

Column 15 gives the average rank for all the final tests, Columns 8 to 13 inclusive. Column 14 gives the average rank for all the tests, Columns 1 to 13 inclusive. From Column 3 (the 30th Vocabulary Test) on, the first four and the last four subjects hold their positions with a surprising degree of constancy. The middle five subjects shift around most. It may not be out of place here to call attention to the achievement of the twins.[10] Peter is seen to maintain his position as

[10]The attention of the reader is further called to the brief sketches of the subjects to be found in the appendix.

the fourth most successful subject rather consistently, but his twin, Paul, exhibits quite an irregular performance which gives him an average rank of ninth.

The following rank correlations were calculated to ascertain the predictability of early vocabulary tests. Rho is uncorrected.

Final Vocabulary Test and early records	rho = .79±.07
Final Vocabulary Test and Vocabulary Test 30	rho = .98±.008
Final Vocabulary Test and Vocabulary Test 40	rho = .97±.01
Final Vocabulary Test and Vocabulary Test 50	rho = .97±.01

As far as these few subjects are concerned, then, the rank in the thirtieth lesson was an excellent prediction of what it would be at the end of the experiment.

The following Pearson product-moment correlations between the scores in the Final Vocabulary Test and the other reading tests are of interest.

Final Vocabulary and Improvised Test	$r = .88 \pm .04$
Final Vocabulary and Wag and Puff Test	$r = .89 \pm .04$
Final Vocabulary and Merrill Tests	$r = .95 \pm .02$
Final Vocabulary and Detroit Test	$r = .86 \pm .05$
Final Vocabulary and Pressey Test	$r = .94 \pm .02$

Special Tests Read by the Two Best Subjects

1. *Puff is Sleepy*. This unstandardized reading test, a rearrangement of the story in the Wag and Puff Primer but containing several new words, was given to Sally and Jean only. It is 104 words in length. In addition to the story, a list of words and phrases was presented for recognition. According to the manual (48), this test is given to first-grade children on the completion of 55 daily lessons. No norms are given, but the manual states, "When he reads he should no-

tice right away that this story is different from the one he read before about the same thing. He will stumble somewhat in his effort to read it as it is and that will be a sign that he is getting from this stage the things you hoped he would get. If the child gives good substitutes for words, that will be to his credit, for it will show that he has the content in mind even though he is word conscious, and that he is able to get words from the context of the story."

Both subjects, in answer to the question, "What is the story about?" read the title correctly. Sally read the story in 2 minutes 15 seconds with five errors, and Jean in 3 minutes 24 seconds with seven errors. Sally read 37 of the 39 words and phrases correctly, and Jean 34. These results must be considered very satisfactory. Both subjects read with a surprising degree of fluency and accuracy.

At the time of the retests, Sally read this story in 53½ seconds with only two errors (omissions). She was reading with extreme rapidity at this time, so fast that she barely took time to enunciate her words clearly. One might almost say that she was reading in a manner half way between oral and silent reading. Further reference will be made to this. Jean read the story in 5 minutes 42 seconds with fifteen errors. She was fatigued before she began, and this is very apparent in the results. They cannot be considered a fair measure of her ability. Sally made no error in the words-and-phrase test, while Jean recognized only 20 of the 39 items.

2. *The Boy Who Ran Away.* This unstandardized reading test was also given to Sally and Jean only.

According to the manual (48), it is given to first-grade children on the completion of the first grade, after they have read both the primer and the first reader. Neither of these subjects had finished reading the primer (much less the reader), Sally having almost completed it, while Jean had read about half. In addition to reading this unfamiliar story, the subjects were asked seven questions about it and were also presented with a list of words for recognition. First-grade children are expected to read the story silently in three minutes, and orally with not more than ten errors. Sixty per cent of children are expected to answer all questions correctly after reading the story silently.

Both children were given three minutes in which to read it silently. They, of course, read it aloud. Sally almost finished reading it within the three minutes, reaching the words *away again* in the third from the last sentence. Jean succeeded in reading about half of it, reaching the words, *from home*. Sally read it orally in 2 minutes 58 seconds—just under the three minutes with ten errors. Jean read it in 4½ minutes with seventeen errors. Sally answered six of the seven questions correctly, and Jean five. As was to be expected, questions six and seven, dealing with the seasons of the year, were beyond their comprehension. Sally's success with question six was probably a lucky guess. Sally recognized 27 and Jean 22 of the 32 words in the word test. First-grade children are expected to recognize only 16, so both subjects were very superior in this respect.

The ability displayed by these subjects in this test must be considered very satisfactory. This was the

longest story they had ever read (it contained 157 words). Many of the words were unknown to them, but they made excellent use of the context. Their comprehension of the story must also be considered excellent. Their lack of knowledge of the seasons of the year can very readily be laid to their age. Sally's performance may well be considered as fulfilling the requirements laid down for average children at the end of the first grade, and she achieved this success in less than half the time taken by average first-grade children, and with much less instruction. Jean's achievement was not as good as Sally's, yet, when everything is taken into consideration, it must be regarded as very creditable.

At the time of the retests, Sally read this story silently in 1 minute 16 seconds, and orally in 1 minute 15 seconds, making one error on both occasions. She answered six of the seven questions correctly, and recognized 30 of the 32 words in the word test. Considering that only two months had elapsed since the final tests, this is a splendid improvement. The rapidity and accuracy of her reading were remarkable.

The experimenter was very reluctant to give Jean this test as she was very fatigued. Towards the end she began to misname words previously recognized correctly. She took 8½ minutes and made 37 errors. Her comprehension was excellent, however, for she improved her score by one point, answering six questions correctly. She was not given the word test. This performance is far from doing her justice.

Special Tests for One Subject

At the close of the retesting period Sally was given

two additional tests, viz., Haggerty Reading Examination, Sigma I, and Gates' Primary Reading Tests.[11]

Haggerty Reading Test. This test is made up of two parts: Test I tests the reading of directions, while Test II consists of 20 questions to be answered by Yes or No. The score of Test I is the number of directions correctly carried out, and Test II is scored "rights minus wrongs." Twenty minutes are allowed for Test I and two minutes for Test II.

Sally obtained a score of twelve in Test I and one of eight in Test II. Both scores are equivalent to the norms for Grade 2 and to a reading age of eight years. According to this test, then, this subject was reading as well as a normal eight-year-old child who has finished the second grade. She was barely three years and seven months old at this time. She attacked the test with the utmost confidence and read with great rapidity. She read the first four pages of directions with but one error in 12 minutes and 5 seconds when she announced that it was too hard for her and she was permitted to stop. She only carried out the directions, however, for the first two pages, all of which were correct. The third and fourth pages were apparently too difficult for her to grasp. It is believed that, if she had been given a rest before beginning the third page, she would have succeeded in carrying out at least some of the directions. This type of test demands considerable concentration, and a child as young as this becomes quickly fatigued if too much of this is demanded at any

[11]The Gates' Primary Reading Tests were not published until several months after the close of our first experiment, and so were not available for the final tests.

one time. Ordinary reading of stories is much less exacting. This subject had never before read to the experimenter for so long a period.

Test II gave her no trouble. She finished 12 out of the 20 questions in the two minutes allowed, making only two errors. It would be entirely fair to give her a score of ten in this test instead of one of eight, as was done, since the method of marking "rights" minus "wrongs" is resorted to in order to penalize guessing. In this case these ten questions were correctly read and answered.

Gates' Primary Reading Tests. These tests, like the Detroit Word Recognition Test, make use of pictures, but their arrangement is much more satisfactory. They consist of three parts: Type 1 tests word recognition, Type 2 tests word, phrase, and sentence reading, and Type 3 tests the reading of directions. Fifteen minutes each are allowed for Types 1 and 2, and 20 minutes for Type 3. The score for Types 1 and 2 is the number of exercises marked correctly minus one-third the number incorrect, while the score for Type 3 is the number of directions correctly followed. All three types were given on the same morning with intervals of play and rest. This was indeed a very heavy program for this young subject.

In Type 1 Sally obtained a score of 31 1/3 in 10 minutes 35 seconds. This is equivalent to a Reading Grade[12] of 2.5 and a Reading Age of 8.0. In Type 2

[12]A very finely graded scale of norms is presented for these tests. Each age and school grade is divided into ten steps and norms given for each step. A Reading Grade of 2.5 means ability to read like an average child who has successfully completed five-tenths of the work of Grade 2, i.e., Grade 2B. A Reading Age of 8.0 means ability to read like an average child of eight years of age.

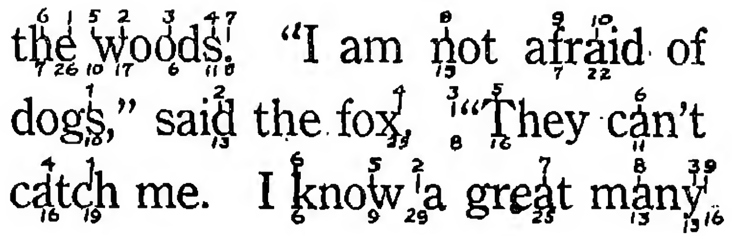
she obtained a score of 17 1/3 in 13 minutes 59 seconds. This is equivalent to a Reading Grade of 2.3 and a Reading Age of 7.8. In Type 3 she obtained a score of 15 in 17 minutes 19 seconds (included in this time is a rest period of four minutes). This is equivalent to a Reading Grade of 2.1 and a Reading Age of 7.6. When the average of these scores is calculated, Sally is found to have a Reading Grade of 2.3 and a Reading Age of 7.8, that is, she was reading at this time as well as an average child not quite eight years old who is in the latter half of Grade 2B.

All three types required careful reading, the Directions Test, in particular, demanding a high degree of concentration. It proved fatiguing for this young subject, so usually insatiable, despite the rest periods provided. After working absorbedly for 10½ minutes in the Directions Test, she declared she was tired. Since she had completed 22 out of the 26 directions in half the time allowed for the whole test, the experimenter felt justified in permitting her to rest for a time. At the end of four minutes the subject resumed the test, finishing it with nearly three minutes to spare even when the rest period is included in her total time. In reality, she took only 14½ minutes in place of the 20 minutes allotted first- and second-grade children for this test.

A Photographic Record of Eye-Movements While Reading

At the close of the retesting period, through the kindness of Dr. W. R. Miles, Professor of Experimental Psychology at Stanford, Sally's eye-movements

were photographed while she read an unfamiliar story. In order to familiarize her with the apparatus,[13] she was first allowed to read a familiar story, and then she was given an entirely unfamiliar story[14] from which the following record[15] was obtained. The figures above the words refer to the order and position of the fixations, while those immediately below indicate the duration of the fixations expressed in fiftieths of a second.



Sally made 10 fixations in the first line, 6 in the second, and 9 in the third, or an average of 8.3 fixations per line. The unfamiliar word *woods* in the first line caused her difficulty, but she seemingly read the rest with great ease, grasping at times two, and even three words in a single fixation. The third line is of special interest, for she apparently first scanned the whole line to see what it was about, and then returned to the beginning of the line and read it. Her reading was also

[13]For a description and diagram of this apparatus see the following article by Miles and Shen (110).

[14]A copy of this story, together with the questions asked about it, will be found in the appendix.

[15]This was the only section of the record that proved suitable for analysis, most of the rest being spoiled by head movements, for it was felt inadvisable to clamp this subject's head because of her age.

singularly free from regressive movements. She made 2 in the first line, 1 in the second, and 2 in the third, or an average of 1.7 regressive movements per line. The average duration of her fixations is 14.7 fiftieths of a second for the first line, 13.83 for the second, and 16.2 for the third, or a grand average of 14.9 fiftieths or 7.5 twenty-fifths of a second for all three lines.

It is extremely interesting to compare these records with the norms presented by Buswell (87). At first the writer compared the number of fixations made by this subject with Buswell's norms for oral reading. Her average score of 8.3 fixations was found to be equivalent to the norm for high school sophomores. As this was so unexpectedly high, an explanation was sought. The following possibly accounts for this situation. Since the close of the experiment, when she was reading with delightful expression and clear enunciation, until the time of the retests when this record was made, she had read a great deal by herself at home. Reading for the enjoyment of the story and with no audience, she had developed a very rapid type of reading, half under her breath—so rapid that she did not always finish the word she had begun to pronounce. This might well be considered a compromise between oral and silent reading. In view of this, it was thought advisable to compare her records with the silent, rather than with the oral, reading norms. On the other hand, of course, it may be that Buswell's norms, which are standardized on only a few subjects in each grade, are not entirely reliable. The following table is taken from Buswell (87).

It will be seen that Sally's scores in all three meas-

TABLE 11

GRADE MEDIANS FOR EYE-MOVEMENTS IN SILENT READING
(From Buswell)

Grade	IB	IA	II	III	IV	V	VI	VII	Fr.	So.	Jr.	Sen.	Col.
Av. no. of fixations per line	18.6	15.5	10.7	8.9	7.3	6.9	7.3	6.8	7.2	5.8	5.5	6.4	5.9
*Av. duration of fixations	16.5	10.8	9.1	7.9	6.7	6.3	5.9	6.0	6.1	6.2	5.6	6.2	6.3
Av. no. of regressive movements per line	5.1	4.0	2.3	1.8	1.4	1.3	1.6	1.5	1.0	0.7	0.7	0.7	0.5

*Expressed in twenty-fifths of a second.

urements lie between the medians for third and fourth grade. From this it may reasonably be concluded that, at least in some cases, it is possible to establish satisfactory eye-movements while reading at the early age of three years and seven months.

Not only did she read the story with considerable ease, but she understood what she read. The writer had prepared eight questions on the content of the story. As soon as the subject had finished reading it, she was asked these questions, the writer recording the answers in a notebook. She answered five and one-half questions correctly. That she did not answer all eight of them correctly may have been due to the excitement of the occasion—the dark room, the strange apparatus, the uncomfortable position in which she had to sit, and so on.

A MOVING-PICTURE RECORD OF EYE-MOVEMENTS WHILE READING

Dr. Miles also took a moving picture of her eyes

while she read another unfamiliar story. This was done in the following manner. The subject was seated out-of-doors, but within the shade, before a flat board tilted at a slight angle. In the middle of the board was a hole through which the lens of the camera projected. The reading material was placed directly above and below the camera lens. At a given signal the subject began to read while Dr. Miles cranked the camera. A record of the eye-movements, at fairly close range, was thus obtained. In it the eyes were found to move in an entirely normal manner.

V

HOW WORDS ARE PERCEIVED IN LEARNING TO READ

The subject of this chapter is of such importance to primary reading that it deserves searching investigation. The present study possibly throws a little light on it in connection with the results of the vocabulary tests, and the data will be given for what they are worth.

All errors made in the vocabulary tests, including the final but not the retest, were analyzed in an effort to find out why the word given had been confused with the word exposed. As many categories were made as there were different types of errors. No category is given in the table below unless it contains ten or more cases, with the exception of those that are related to the larger classifications. No word is placed in more than one category. There were 572 words in all. It was no easy matter, in many cases, to know into which category a word should be placed, as it could obviously fit more than one. The writer attempted to choose what seemed to her the most reasonable category, and was assisted by her observations of the subjects while they matched words, or by remarks they made at that time. The writer is well aware, however, that others might not agree with all her decisions.

Table 12 presents some very striking features. The largest single category is that in which the geometric shape of the word appeared to be the dominating factor. This contains 89 words, or 15.6% of the 572

TABLE 12

ANALYSIS OF ERRORS

The following points of identity were found to exist between the word wrongly given and the true or exposed word.

A.	Geometric shape	89
	Guesses	78
	Peculiar letter	40
	Associated in context	28
	Derivatives	14
	Names of colors confused	12
	Middle letter identical	10
	Other categories	26
	Total	297
B.	First letter	59
	First two letters	18
	First three letters	8
	First and one other letter	8
	First and two other letters	8
	First and three other letters	2
	Total	103
C.	Last letter	29
	Last two letters	21
	Last three letters	4
	Last four letters	1
	Last and one other letter	4
	Last and three other letters	1
	Total	60
D.	First and last letters	7
	First two and last two letters	1
	First and last and one other letter	8
	First and last and two other letters	1
	First and last three letters	4
	Total	21
E.	Inverted geometric shape	23
	Word begins with last letter of exposed word	20
	Word ends with first letter of exposed word	26
	Word begins with last letter of exposed word and contains one other identical letter	13
	Word begins with last letter of exposed word and contains two other identical letters	2
	Word begins with last two letters of exposed word but in reverse order	7
	Total	91
	Grand total	572

causing 5.1% of the total errors. Meek (111), too, found that the initial letter was more often used as a cue than the final letter. She found, also, that the last two letters acted as cues more frequently than the first two. There is only slight substantiation of this fact in the above findings.

An interesting category is that called "associated in context." This is the sixth largest one, containing 28 words. In this category were placed words that were introduced in the same lesson, or appeared frequently in the same story, and which could not seemingly be explained in any other way. Adults make such confusions in many life situations. Most of the other categories are self-explanatory.

It will be noted in Table 12 that the categories have been grouped in larger classifications, A, B, C, etc. With the exception of A, the categories within each group are obviously related. It will be seen at once that the first part of the word is used more often as a cue than the last part, the former occurring 103 times to 60 times for the latter, or, expressed in percentage of the total 572 words, 18% to 10%.

The writer wishes to draw particular attention to classification E. This is the most striking feature of the whole table and deserves special consideration.

The writer had noted on different occasions that the subjects frequently gave a word that began with the last letter of the word exposed, e.g., *table* instead of *rabbit*. Analysis of all errors confirmed this observation. The above happened in 20 cases. In 26 cases, the word given ended in the same letter as formed the beginning of the exposed word, e.g., *dog* for *green, boy* for *yes,*

cow for *was.* In 13 cases, the word given began with the last letter of the exposed word, but, in addition, there was one other letter identical to one in the exposed word. In seven cases, the first two letters of the given word were identical with the last two in the exposed word, but in reverse order, e.g., *draw* and *warm, boy* and *you, not* and *to, car* and *ran.* In 23 cases, not only did the word begin with the last letter of the exposed word, or vice versa, but when the word was inverted, there was a distinct geometric likeness to it. All 91 cases in category E—15.9% of the 572 cases—bear a resemblance to each other. They have suggested to the writer the possibility that left-to-right inversion of words[10] is no more confusing to young children than is upside down inversion. By this is meant that, to young children, words, like toys, do not have a right and a left side; that children can read them almost as easily from right to left as from left to right. Bowden (86) noted that children can read words upside down as easily as right side up, as did the writer, but apparently no English or American writer has noted this matter

[10]Since the present thesis was completed in manuscript form, S. T. Orton has published his neurological theory of reading disability in "An impediment to learning to read—a neurological explanation of the reading disability," which is printed in *School & Soc.,* 1928, 28, 286-290. He bases his theory on the fact that children (*a*) confuse letters of the same form but facing differently, (*b*) read from right to left instead of from left to right, and (*c*) read mirror writing with ease. He considers reading disability as due to "a lack of establishment of clear cut dominance in one hemisphere." He also briefly expounds his theory in his introduction to "Methods for Diagnosis and Treatment of Cases of Reading Disability" by Marion Monroe, which is published in *Genet. Psychol. Monog.,* 1928, 4, 343-456.

of left-to-right inversion of words in beginning reading. It does not seem unreasonable in view of certain well-known facts. It is a matter of everyday observation that children can recognize pictures in their picture books at various angles. This had led certain psychologists to express the belief that form to young children is independent of absolute spatial position. Koffka (52) comes nearest to corroboration of the findings of this experiment when, after discussing the fact that children can recognize pictures in various positions, he states: "This peculiarity continues for a long time. Even at the beginning of the school-period it may be noticed that many children copy the letters given them, not only in the right position but in all possible positions; as for instance in mirror-writing, or upside down. Teachers who at my request have made observations upon this subject, have reported that certain children can read mirror-writing at first just as well as ordinary writing, which shows the difference between children and adults, for an adult finds it no easy task to read mirror-writing. Originally then a figure is in a high degree independent of its position, whereas for adults the absolute orientation of the figure is a very powerful factor. Right and left, above and below, become characteristic properties of the different members of the configuration; and consequently of the total-form."

This indication that left-to-right inversion of words causes little confusion to very young children would seemingly confirm the aforementioned belief that form is independent of its absolute spatial position for young children, and it implies, at the same time, that words

possess a very definite geometric shape. Because of the light it would throw on the perception of words in particular, and on the perception of form in general, the writer agrees with Koffka that "A closer investigation of the development of this positional factor in children's perception would certainly prove a stimulating and valuable undertaking."

VI

FACTORS RELATING TO READING SUCCESS

This survey of the results of the reading tests has shown that the bright three-year-old children were superior to the average four- and dull five-year-old children. It has also revealed the presence of large individual differences. Whether these are due to mental-age differences, intelligence-quotient differences, or to other factors must next be considered. Analysis of the Stanford Binet Test results should throw light on the first two factors mentioned.

Factors Revealed by the Stanford Binet Test Results

Each subject received three Stanford Binet Tests, one at the beginning of the experiment on the basis of which the subject was selected, one at its close, and one at the time of the retests. These will be referred to as Binet 1, Binet 2, and Binet 3, respectively. Binet 1 will be considered first.

According to the chronological ages given in Table 2, Group A was slightly more than one year older than Group B, and Group D was ten months older than Group A. Group B had the smallest age range. The most successful subject in reading was the youngest one, this subject having been tested two days after her third birthday, while the least successful was the second oldest. A Pearson product-moment correlation between chronological age and success in reading, as

measured by the Final Vocabulary Test, proved to be $-.56 \pm .13$. This means that the younger the child the more successful he was in learning to read, as was indeed the case. This, of course, would not hold true of unselected children.

As already stated, the subjects were chosen so that their mental ages would fall between four and four and one-half years. When it was found impossible to secure sufficient subjects within these age limits, exceptions were made, but an effort was put forth to match mental age with mental age as far as the material at our disposal permitted. From Table 13 it is seen that Group A had the highest average mental age and Group D the lowest, the latter group having also the smallest mental-age range. Group B, however, was the most successful group in reading. The mental age of the most successful subject was 4-2. This mental age was also gained by Donald of D who was the least successful subject, by Edgar of D, Dora of A, and Kate of B. These subjects displayed quite a wide range of success. The subjects with the lowest mental age, namely, 3-8, were Bell of D and Leo of A, both of whom did rather poorly.

According to the same table, the average IQ of Group B was 29.5 points higher than that of Group A, and the average IQ of Group A was 21.5 points higher than that of Group D. It is further seen that the IQ's of the groups did not overlap. These figures indicate that the groups were distinct as regards "brightness." The subjects were selected so that this would be the case. The two most successful subjects had the two highest intelligence quotients, while the least successful

subject had the third lowest one. The Pearson product-moment correlation between Final Vocabulary score and intelligence quotient was .69±.10. (The correlation between Final Vocabulary score and the average intelligence quotient for all three tests was .70±.09.) This bears out the correlation with chronological age, for the youngest children were deliberately chosen to be the brightest children. "Brightness," then, is apparently the most important factor in reading success so far considered.

TABLE 13

Comparison of Three Stanford Binet Tests

	Mental ages			Gains in MA		
Binet	1	2	3	2-1	3-1	3-2
Bright						
Jean	4-6	4-7	4-3	1	—3	—4
Paul	3-10	4-2	3-10	4	0	—4
Peter	4-0	4-5	3-10	5	—2	—7
Kate	4-2	4-3	3-10	1	—4	—5
Sally	4-2	4-8	4-7	6	5	—1
Av.	4-1.6	4-5.0	4-0.8	3.4	—0.8	—4.2
Grand Av. of 1, 2, 3	4-2.5					
Average						
Leo	3-8	3-11	3-11	3	3	0
Elsa	4-6	4-10	4-11	4	5	1
Lela	4-6	4-10	5-0	4	6	2
Dora	4-2	4-10	4-6	8	4	—4
Av.	4-2.5	4-7.3	4-7	4.75	4.5	—0.25
Grand Av. of 1, 2, 3	4-5.6					
Dull						
George	3-10	4-2	4-2	4	4	0
Edgar	4-2	3-10	—	—4	—	—
Donald	4-2	5-7	5-9	17	19	2
Bell	3-8	4-6	4-8	10	12	2
Av.	3-11.5	4-6.25	4-10.3	6.75	11.7	1.3
Grand Av. of 1, 2, 3	4-5.35					

Av. gain in MA (Binet 2) for all three groups = 4.97 mos.

TABLE 13 (*continued*)

	Intelligence quotients			Gains in IQ		
Binet	1	2	3	2-1	3-1	3-2
Bright						
Jean	138	141	130	3	—8	—11
Paul	115	124	115	9	0	—9
Peter	120	133	115	13	—5	—18
Kate	128	132	117	4	—11	—15
Sally	139	156	153	17	14	—3
Av.	128	137.2	126	9.2	—2.0	—11.2
Grand Av. of 1, 2, 3	130.4					
Average						
Leo	88	95	95	7	7	0
Elsa	100	107	110	7	10	3
Lela	104	112	115	8	11	3
Dora	102	119	110	17	8	—9
Av.	98.5	108.25	107.5	9.75	9.0	—0.75
Grand Av. of 1, 2,3	104.75					
Dull						
George	77	83	84	6	7	1
Edgar	76	70	—	—6	—	—
Donald	82	109	113	27	31	4
Bell	73	89	93	16	20	4
Av.	77	87.75	96.6	10.75	19.3	3.0
Grand Av. of 1, 2, 3	87.12					

Av. gain in IQ (Binet 2) for all three groups =9.90 points.

That subjects with approximately the same mental age should show such a variety of success in reading may be because the mental ages obtained by a single Stanford Binet are not reliably measured in these early preschool years. This is commonly believed. It is not unreasonable to imagine that the novelty of the situation, the strange surroundings, and the strange experimenter should prevent these youthful subjects from doing their best, despite the care taken to minimize these as far as possible. It is a fairly simple matter to

fail in one or two tests, but each failure in a test means a loss of two months, and a loss of two months in these early years is much more serious than at eight or nine years of age. Fortunately, every subject but one had received two additional Binet tests. As a consequence, it is possible to compare these tests and to determine, to some extent, their reliability. In order to make all mental ages comparable, it was necessary to calculate them from the same point of time; this was done. The mental ages presented are based on the IQ's obtained in Binet 2 and 3, and on the CA's in Binet 1.

All three groups made gains in Binet 2. Group B made the smallest and Group D the largest gains. The respective gains in MA for Groups B, A, and D, were 3.4, 4.75, and 6.75 months, and in IQ were 9.2, 9.75, and 10.75 points, about twice the probable error of the score. Group B was now the only group to fall within the limits set up for the experiment, the other two groups being slightly in excess of them. Group D now showed the widest range of mental age, its upper limit of 5-7 being gained by the least successful subject. This subject's behavior in activities other than reading led the writer to believe that Binet 1 underrated him. This new mental age is believed to overrate him, a belief confirmed by a fourth Binet given two years later. Some of his gains were no doubt aided by school experience, but some were no less due to willingness to put forth effort, a very undependable factor where this subject was concerned.

It is interesting to compare these gains in IQ with those obtained by Goodenough (42) for 28 nursery-school children tested with the Kuhlman Binet before

and after one year's nursery-school experience. Her younger group made an average gain of 9.9 points in IQ—the exact average gain of our three groups—but her older group made a larger gain than any of our groups, namely, 12.5 points. (Her control groups gained 8.0 and 8.8 points, respectively.) She obtained changes in IQ for all 28 nursery-school children ranging from —4 to +27 while those obtained in this experiment were from —6 to +27.

Dr. Goodenough does not give the correlation obtaining between these tests for this group of children, but she found a correlation of .813±.012 between the first and second IQ's of the 300 children of her main experimental group. The Pearson product-moment correlation between the IQ's of Binet 1 and Binet 2 of this experiment was found to be .95±.02, a very substantial correlation, and very similar to that obtained for older unselected children. [See the summary in the Twenty-Seventh Yearbook, Part I (59).]

Dr. Goodenough attributes the changes in IQ to irregular standardization of the Kuhlman Binet at the early years rather than to nursery-school experience. Whether this can equally well be said of the gains in the present study cannot be determined without a comparable control group. That the Stanford Binet is imperfectly standardized in these early years is generally acknowledged.[17]

A correlation between the Final Vocabulary Test and gain in IQ was found to be —.026±.18.

---

[17]An extensive restandardization of the Stanford Binet Test, with special attention directed to the preschool and high school ages, is being carried out at Stanford University at the present time.

In Binet 3, Group D was the only group to make a still further gain, a slight one based on only three subjects; Group A made little change, and Group B lost more than it had gained in Binet 2. Group D's gain was again due to the large individual gains of Donald and Bell. A fourth Binet, given two years after Binet 1, showed the former to have an IQ of 102, a much more reasonable measure of that subject's intelligence. Bell's IQ of 93 seems rather high in view of her behavior. Dora's IQ dropped from 119 to 110. A fourth Binet rated her at 108 IQ. Binet 3 is perhaps a better estimate of Sally's ability than is Binet 1, but it does not do Jean nor Kate justice, and the others of this group may be somewhat underestimated.

The correlation between the IQ's of Binet 1 and Binet 3 was found to be $.86 \pm .05$ and that between Binet 2 and Binet 3, $.95 \pm .02$. The latter is the same as that found between Binet 1 and Binet 2.

The fact that in both retests the young bright three-year-old children made the smallest gains raises the interesting question as to whether their home environment was more stimulating than that of the other two groups and that, as a consequence, this led to a fuller and earlier development of their latent capacities, while the others had to wait until exposed to school conditions to enjoy a similar development. It would be interesting to know, also, whether the gains made will prove to be permanent.

Developmental and Environmental Factors

It has now been seen that when mental age is kept fairly constant the most important factor so far con-

sidered in success in learning to read is "brightness." It is now necessary to consider the possibility of the presence of other factors that might contribute to this success. In order to throw light on this question the parents were asked to fill out information blanks covering the following main points:

1. Information concerning child
   a. Developmental history (age on getting first tooth, first walking, first talking, speech difficulties).
   b. Evidences of precocity or backwardness.
   c. Child's interest in play and in stories.
   d. Frequency of being read to.
2. Information concerning parents
   a. Occupational status of fathers.
   b. Education of both parents.
   c. Reading and recreational interests of both parents.

Developmental factors. The data on developmental history are presented in Table 14. Neither the age on acquiring the first tooth nor the age at beginning to walk bears any obvious relation to ability to learn to read. The means for the three groups reveal no striking differences, nor does the range reveal any abnormalities when compared with the range for the Stanford gifted children (71). Age on first beginning to talk, on the other hand, might conceivably bear some relation to learning to read, since reading involves speech. The results of the Stanford study of gifted children (71) and Mead's norms (54) are given below for comparative purposes.

(Figures refer to age in months)

	First tooth	First walking	First talking
Stanford gifted girls	7.02	12.87	11.01
Stanford gifted boys	7.18	13.10	11.74
Mead's normal group		13.88	15.32
Mead's feebleminded group		25.02	38.52

The average ages at which Groups B, A, and D began to talk were 13, 14, and 20.25 months, respectively. Groups B and A are thus seen to have been very much alike in this respect, but Group D was 7.25 months older than Group B. This seems to be a striking difference. A correlation between age at beginning to talk and success in reading as measured by the Final Vocabulary Test, however, was found to be $-.14 \pm .18$ which is not statistically significant.

The last column of Table 14 deals with speech difficulties. In several instances the statements made by the parents did not agree with the writer's observation of the subjects. This was especially true of Group D. Group B used less "baby talk," mutilated words less, and enunciated more clearly than either of the other groups. Group D was the poorest in these respects, Donald being the only distinct speaker in this group. Both George and Edgar mutilated words to a considerable extent, while it was very difficult to understand Bell at all. Kate of Group B showed considerable precocity in language development.

That primogeniture does not apparently play a part in reading success is seen from the following. (Figures in parentheses after the name refer to the number of children in the family besides the subject.) Sally and

good memories for either poems, music, or names of places and persons: Jean, Sally, Peter, and Kate of B; Elsa, Lela, and Dora of A; and George of D. The following possessed unusual constructive ability: Paul and Kate of B; Donald and Edgar of D. The following were strongly imaginative: Jean, Paul, and Kate of B; Elsa, Leo, and Lela of A; and Donald of D. Paul of B showed unusual ability in reasoning. During the experiment a few cases of outstanding ability were noted by the writer. Jean of B showed a remarkable ability to memorize poems and songs. Donald of

TABLE 15
FREQUENCY OF BEING READ TO BY PARENTS
(* indicates the category into which each subject fell)

	Never	Occasionally	Frequently	Regularly 2 or 3 times per week	Regularly every day	Favorite stories
Jean					*	Animal stories; real life; early childhood of parents
Paul				*		Little boys; animals
Peter				*		All kinds
Sally					*	Fairies, children, animals; no stories with unhappy incidents nor about foxes
Kate					*	Poetry
Total				2	3	
Leo		*				Animal
Elsa					*	Fairy; comic sheet in newspaper
Lela		*				Animal and adventure
Dora		*				Animal
Total		3			1	
George		*				Stories in his primer
Edgar		*				
Donald		*				Child's life of Christ
Bell		*				*Not particular*
Total		4				

D and Kate of B exhibited an absorbing passion for constructing. Undoubtedly this enthusiasm for building acted as a strong counter-attraction to the reading situation.

Environmental factors. Table 15 shows with what frequency the children were read to by their parents. It was thought that those children in whom an early love for books and stories was inculcated in the home might be more easily interested in learning to read than those who did not have this stimulation. This table is consequently rather striking. It divides the subjects clearly into two groups, those read to regularly and those read to occasionally. No one falls into the middle group. All the subjects of Group B are in one category and all but one of Groups A and D are in the other category. This subject, Elsa, of A, was the third most successful subject. From this it is seen that the subjects who were the most successful in learning to read were the subjects who were read to most frequently. But these were the subjects who had the highest IQ's. Consequently, it is impossible from the data presented to determine whether the success was due to the stimulating environment or to their having parents of superior intelligence, who, because of this superior intelligence would naturally be more likely to read to their children. It does not seem unreasonable to believe, however, that the stimulating environment at least helped to create an attitude towards books and stories favorable for the arousal of a desire to read. Of course, on the other hand, it is not impossible to conceive that a child who has an adult ever ready to

satisfy his demands for a story might find learning to read too tedious a task.

It was further considered that the parents' own attitude towards books might very well influence that of their children. Considerable data on the amount and quality of the reading done by both parents of these subjects were accordingly obtained. It is not necessary to give the details here. Suffice it to say that both in the amount and the quality of the reading done, the parents of Group B were unmistakably superior to the parents of the other two groups, between whom there was little difference.

Two other important environmental factors were studied, namely, the occupational status of the fathers and the education of both parents. These will now be considered.

According to the Taussig (70) scale, all four fathers of Group B are in class five, the highest category; all eight fathers of Groups A and D are in class four. Three fathers in each of Groups A and D are in business for themselves. Three fathers of Group B are professional men and the fourth holds a high executive position in the business world.

When the educational advantages enjoyed by the parents were analyzed, it was found that the parents of the bright subjects had received much better educations than the parents of the other groups. All the parents of Group B but one mother held university degrees, and this mother had attended a university for two years. No parent in either of the other groups had received a university education. The mothers of Group A had received a somewhat superior education

to those of Group D. Two of the former had completed a two-year normal school course, one had completed high school, and one had attended high school for two years. None of the mothers of Group D had completed high school. One had attended high school for two years, two for one year, and the fourth had completed the elementary school only. The fathers of Groups A and D had somewhat similar educations. The fathers of Group A had completed the elementary school, and one had attended it for seven years. Later on, however, this last parent had taken a four-year university extension course in sanitary engineering, and, at the time of the experiment, had a business of his own. One father of Group D had completed high school, while the others had completed the elementary school, one of the latter, in addition, having received a business college education. It is possible that this superiority in the education of Group B reflects their superior natural intelligence.

In addition to the factors already discussed several other possible factors remain to be mentioned. It is possible that visual imagery plays a part in the reading situation. No data were obtained on this point. They are difficult to secure with adults and much more so with young children. Another possible factor is that of memory span. This was not measured adequately, the only data available being the incomplete records of the memory span for digits gained from the Stanford Binet tests. One subject, Jean of Group B, passed the six-digits, but failed in the seven-digits test, in her third Binet, the only time when she was fully tested, that is, tested until she failed. Since this test consists

of only three trials with each series it can have only a very low reliability, but, in this instance, the result confirmed the writer's observation of this subject's remarkable ability to memorize lengthy poems. Sally, the most successful subject, was only able to repeat four digits, while Donald, the least successful subject, could repeat five. One further possible factor is that of a special aptitude for reading. Whether or not this is merely a happy combination of all the factors previously mentioned, together with a method that happens to be particularly suited to display it to the best advantage, is difficult to determine. It may be merely a case of success breeding success, for there is little doubt but that success is an important factor in maintaining interest. Finally, there is a possibility that an overpowering interest in some other activity may act as a counter-attraction to the reading situation, and delay its development in these early years.

Summarizing briefly the main facts brought out by this discussion, it is seen that the subjects of Group B enjoyed the following environmental advantages over the subjects of the other two groups: their fathers held a higher occupational status; their parents had received superior educational advantages; they read a better type of book and not only read more themselves but read more regularly to their children than did the parents of the other subjects. The data on precocity and backwardness displayed by the subjects shed no light on reading success, nor did age at first beginning to talk, speech difficulties, nor primogeniture. No data, or inadequate data, made it impossible to determine the relationship, if any, between visual imagery

or memory span and success in reading. In short, the *group that attained the greatest success in reading in* this experiment was the group that had gained the highest intelligence quotient in the Stanford Binet Test and had enjoyed environmental advantages superior to those enjoyed by the other two groups. The data are insufficient to determine whether intelligence or environment is the more influential factor in producing this success. It may be said, however, that so far *as this experiment is concerned, brightness is more im*portant for reading success than chronological age when mental age is restricted within narrow limits, and that an environment which stimulates an interest in books and stories creates conditions probably favorable to the development of reading success.

VII

SUBSEQUENT PROGRESS OF SUBJECTS

No systematic "follow-up" of the subjects of this experiment has been made. At the time of the retests the parents filled out information blanks in which they stated, among other items, to what extent and in what ways the children had benefited from the experiment. The writer heard occasionally from most of the parents who had left town, and, in June, 1928, she visited one of the schools in the neighborhood, where she found four former subjects—three of the average and one of the dull group—in the first grade. She gave these subjects several standardized reading tests, the results of which will be presented together with other available information.

The Bright Group. None of these subjects is as yet old enough to attend school, so information about them is somewhat limited. Their already keen interest in pictures, books, and stories had been further stimulated by the classwork. Information received nearly a year after the close of the experiment stated that Jean often read to the neighborhood children, who "look round-eyed at her to think that she really knows what the words are." Nothing is known about Paul and Peter except that they still talk of the "school" with pleasure. Kate has attended a kindergarten during the past year and has made excellent progress in every way. She is now beginning to express a desire to learn to read, having forgotten in the interval most of what she had learned. Sally's interest in reading has never abated

and, when last heard from, she was able to read from a Fourth Reader while but four years of age.

The Average Group. Two years after the close of this experiment in June, 1926, three of these subjects, Elsa, Lela, and Dora (and Donald of D), were completing first grade, and all were to be promoted to second grade. The teacher reported that Elsa and Lela were "A" students but were doing only "B" work in reading. In view of the ability displayed by these two subjects during the experiment, the writer decided to check the teacher's statements by means of standardized reading tests, the results of which (together with the norms) are given below. (It should be mentioned that this first-grade class was a small and particularly good one, and the teacher was unusually successful in obtaining good results in reading.)

TABLE 16

"Follow-Up" Reading Tests

	Pressey Test Word	Pressey Test Sentence	Pressey Test Total	Detroit Word Recognition	Haggerty Test 1	Gates' Primary I	Gates' Primary II	Gates' Primary III
Elsa	25	15	40	26	8	37	18	14
Lela	25	15	40	21	5	39	23	17
Dora	19	14	33	19	2	0	8	6
Donald	14	11	25	12	0	1	2	6

	Corresponding grade norms (Gates) I	II	III	Corresponding age norms (Gates) I	II	III	Av. norms (Gates) Av. grade	Av. age
Elsa	2.90	2.25	2.0	8.50	7.75	7.50	2.38	7.9
Lela	3.10	2.50	2.3	8.80	8.00	7.80	2.63	8.20
Dora	0	1.60	1.55	0	7.10	7.05	1.55	7.05
Donald	1.20	1.30	1.55	6.60	6.70	7.05	1.35	6.78

The norms for the other tests are as follows:

Grade	Word	Pressey Sentence	Total	Detroit Word Recognition	Haggerty Test 1
1B	9.5	8.5	17.0	12	—
1A	18.5	13.0	31.0	20	4
2B	—	—	—	28	—
2A	—	—	—	—	12

Elsa (CA 6-11; MA 6-11).[18] Rated by her teacher as an "A" student in thought and "B" in reading; spelling equal to second grade. This subject's score in the Pressey test (a perfect one) was nine points above the norm for the end of first grade. Her score in the Detroit test was two points below the norm for the end of Grade 2B. Her score in Test 1 of the Haggerty test (Test 2 was inadvertently omitted) was halfway between the norms for first and second grades. In the three types of the Gates' Primary Reading Tests she obtained Reading Grades of 2.90, 2.25, and 2.0, and Reading Ages of 8.50, 7.75, and 7.50, respectively. According to the average results of the Gates' tests, then, this subject was reading as well as a normal child of 7.9 years of age who is doing the work of Grade 2.38, that is, who is well advanced in Grade 2B. As she was barely 6 years 11 months old at this time, she was thus reading as well as a normal child one year older than herself. This must be considered a very superior quality of work.

[18]The CA and MA of each subject for June, 1928, are given in parentheses after each name, the MA being calculated on the basis of the IQ gained in Binet 1. The MA's for Donald and Dora were obtained from Binets given at this time.

Lela (CA 6-9; MA 7-0). Rated by her teacher as an "A" student but only "B" in reading ability. In the Pressey, Detroit, and Haggerty tests, respectively, this subject gained scores nine, one, and one point above the norms for Grade 1A. In the three types of the Gates' tests, however, she gained Reading Grades of 3.10, 2.50, and 2.30, and Reading Ages of 8.80, 8.00, and 7.80, respectively. According to the average of these, she was reading as well as a normal child of 8.20 years of age who is doing work of the level of Grade 2.63. This means that she was able to read as well as a normal child one and one-half years older than herself. As in the case of Elsa, this subject's reading was of a very superior quality.

Dora (CA 6-6; MA 7-0). Rated by her teacher as a "C" pupil, doing "C" work in reading. In the Pressey and Detroit tests, this subject gained scores respectively one point above and one point below the norms for the end of first grade. In the Haggerty test her score was half the norm for first grade. In the three types of the Gates' tests she received Reading Grades of 0, 1.60, and 1.55 and Reading Ages of 0, 7.10, and 7.05, respectively. It will be seen that this subject has made considerable progress since the close of the experiment. She was at this time only six and one-half years of age, successfully completing the first grade. Judged by the Gates' tests, she was reading as well as a normal child six months older than herself who has finished Grade 1B, but, judged by the Pressey and Detroit tests, her ability is equal to that of a child who has finished Grade 1A, as she has—just what one would normally expect from her, for, according to a

fourth Stanford Binet, given to her at this time, her MA was 7-0 and her IQ 108.

Leo (CA 6-9; MA 5-11). This subject moved away shortly after having been promoted to high first grade in January, 1928. His teacher rated him as a "Low B" pupil, doing "C" work in reading. If his "C" rates in proportion to the "B" 's above, it may be considered that he was doing work superior to his mental level, which was only 5 years 11 months in June, 1928, if his first Binet was reliable.

The Dull Group

Donald (CA 7-6; MA 7-8). Rated as a very poor pupil in every respect. He is a very trying disciplinary problem both to his foster parents and his teachers. He runs away, is undependable and untruthful. In the Pressey test, this subject's score was eight points above the norm for Grade 1B, and, in the Detroit test, it equalled that norm. He received a zero score in the Haggerty test, due largely to inattention. In the three types of the Gates' tests he secured Reading Grades of 1.20, 1.30, and 1.55, and Reading Ages of 6.60, 6.70, and 7.05, respectively. Accordingly, his reading ability was about the level of Grade 1B, and of normal children one year younger than himself. A fourth Stanford Binet, given to him at this time, showed him to have a MA of 7-8 and an IQ of 102, a fair estimate of his ability as judged by activities other than reading displayed by him during the experiment. He was the only subject who was reported to have shown no increased interest in books and stories as a result of the experiment.

George. This subject had left town and no fur-

ther particulars outside of the information blank are available concerning him. The parents stated that, as a result of the experiment, he had gained a greater interest in pictures, stories, books, and in reading, occasionally reading to himself at home.

Edgar. This subject entered school in Pasadena in the September following the close of the experiment. He was put in Grade 1B, but in a few days, he was promoted to Grade 1A, and his first report card contained nothing but "A" 's. This is a very promising beginning, especially for a child who, at this time, had a mental age of only 4 years 9 months, if his first Binet was reliable.

Bell. This subject, too, had moved away. The following was reported by her mother: "Bell entered First Grade the 24th of January (1927) and is doing splendid work. When we moved on the 15th of March the teacher had not started the class to read from books. She had been drilling them entirely on sounds and words from the blackboard. On the 16th of March I placed Bell in the —— school. There the teacher had her class reading and nearly through the first reader. Bell has caught up with the class (in less than three months) and is reading as well as the average pupil. I feel this is due to your training a year ago . . ." This is a very interesting report, especially in view of the fact that Bell must be considered a borderline case. At the time of this report she was six years and three months old, and her mental age was four years and six months, if her first Binet was reliable. It will be remembered that she developed a great enthusiasm for reading, although she usually ranked about tenth in

the group. Bell benefited in many ways from the experiment, her mother reporting on a previous occasion, among many other things, that "her ambition to learn and to do" had been aroused.

From the above information, received from 4 to 24 months after the close of the experiment, it is seen that only one of the seven subjects known to be attending school was a failure, and he was the least successful in the experiment. Two years after the close of the experiment, two subjects, Elsa and Lela, were doing very superior work as judged by standardized reading tests, and one, Dora, was doing at least normal work for her mental age, finishing the first grade when only six and one-half years old. One, Leo, is believed to be doing work quite superior to his mental age, as are Bell and Edgar of D, according to teachers' reports. This record of achievement tends to refute a not uncommon belief that early training in reading is injurious to children. Parents' reports of the benefits to the children from the kindergarten work and games carried on during the experiment are rather convincing as to the value of preschool training. The normal four- and the dull five-year-old children were the ones who seemingly derived most benefit, perhaps because they lived in a less stimulating environment than those of Group B.

VIII

SUMMARY AND CONCLUSIONS

Summary

1. This experiment attempted to find out to what extent children with a mental age of four years could learn to read, and whether bright, average, and dull children, all with this mental age, would learn to read equally well under the same experimental conditions.

2. Three groups of preschool children were used: a bright three-, a normal four-, and a dull five-year-old group. All had mental ages of approximately four years. There were five in the bright group, and four each in the average and dull groups, thirteen children in all.

3. The children were selected by means of the Stanford Binet Test. Their parents were white and were born in this country.

4. Each group met as a group, at a separate time, for a period of nearly four and one-half months. The average and dull children met 80 times and the bright 76 times.

5. After about twelve days of preliminary kindergarten work and games, each subject received a daily ten-minute reading lesson, and the group, as a whole, usually received in addition a brief daily group game. The largest number of reading lessons received by any subject was 61.

6. A special reading method with the following main features was developed for this experiment: the introduction of print by means of a series of Massed

Forms, definite training in the use of cues, careful control of difficulties, use of the subject's love of motor activity, use of a picture-word vocabulary, and the avoidance of eye fatigue through the method of introducing the sentence.

7. At regular intervals throughout the experiment, as well as at its close and again some weeks later, complete vocabulary tests and both standardized and unstandardized reading tests were given.

8. Data concerning the early developmental history of the subjects, and concerning such environmental factors as the occupational status, education, and reading interests of the parents, were obtained.

9. The results of the reading tests revealed the great superiority of the bright three-year-old group over the other two groups; the average four-year-old group was slightly superior to the dull five-year-old group. There were large individual differences in each group.

10. At the close of the experiment the bright three-year-old children recognized on the average 129.4 words out of context, the average four-year-olds 55.3 words, and the dull five-year-olds 40.0 words. The number of words recognized by individual subjects ranged from 20 to 269.

11. Despite the fact that the experiment occupied less than one school term (one-half school year), that the highest number of reading lessons received by any one subject was 61, and that the amount of time spent in teaching reading was very much less than that commonly spent in the first grade, nevertheless, the bright three-year-old subjects gained an average score in the

Pressey First Grade Reading Test 4.8 points above the norm for the end of Grade 1B, while two of them gained scores above the norm for Grade 1A.

12. The two most successful subjects revealed ability at the close of the experiment equal to that of normal children who have completed the first grade. One was three years four and one-half months old and the other three years eight months old at this time.

13. At the time of the retests, when three years seven months old, the most successful subject read as well as a normal child of almost eight years of age who is in second grade. Her eye-movements while reading an unfamiliar story were superior to those of children in third grade as judged by a photographic record.

14. The subjects maintained their relative rankings in the reading tests with marked constancy from the thirtieth reading lesson on.

15. All three groups made gains in the second Stanford Binet, but only the dull five-year-old group made a further gain in the third Binet. The bright three-year-old group made the smallest and the dull five-year-old group the largest gains. The least successful subject in reading made the largest individual gain in IQ.

16. Brightness, as measured by the first IQ, correlated with success in reading, as measured by the Final Vocabulary Test, to the extent of $.69 \pm .10$, while chronological age and Final Vocabulary Test correlated to the extent of $-.56 \pm .13$. A zero correlation was found to exist between score in Final Vocabulary Test and the amount of gain in IQ.

17. The bright three-year-old children used less

"baby talk," mutilated words less, and enunciated more distinctly than the other subjects. The dull five-year-old subjects, on the average, were the poorest in these respects. They had also been the most backward in beginning to talk.

18. The bright three-year-old subjects have enjoyed the following environmental advantages over the subjects of the other groups: their fathers occupy a higher occupational status; their parents, both mothers and fathers, have received very superior educational advantages; their parents read more often to them, read more themselves, and read a better type of book than the parents of the other subjects.

19. As a result of the experiment all the subjects but one five-year-old were reported by their parents to have gained an increased interest in pictures, books, and stories. This was greatest for the four- and five-year-old subjects, the three-year-old subjects always having manifested a keen interest in them. Valuable gains in many other directions were reported for all subjects.

20. Of the five bright three-year-old subjects, the two most successful ones have maintained their enthusiasm for reading since the close of the experiment; the others quickly forgot what they had learned. All the normal four-year-old subjects have attended the first grade for a year. According to the results of standardized tests or teachers' reports, three of them are doing work superior to their mental age and one work at least equal to it, for she is being promoted to second grade at the age of six and one-half years. Three of the dull five-year-old subjects are known to be in school. One

of these was doing work inferior to his mental age, and two, when last heard of, were doing work superior to their mental ages. Nothing is known of the fourth.

21. Analysis of errors made in the vocabulary tests revealed the fact that the first part of the word—either the first letter alone or in other combinations—was the most frequently used cue in the recognition of words; next was the geometric shape of the word; and, third, the last letter alone or in combination with other identical letters. Words were also confused when they contained the same oddly shaped letter, had been associated in the same context, were derivatives of known words, or contained the same middle letter.

22. Analysis of errors has revealed the possibility that left-to-right inversion of words caused little confusion to very young children. This would seemingly confirm the belief expressed by several psychologists that to young children form is independent of absolute spatial position. If this should indeed prove to be the case, it would also argue for the definiteness of the geometric form of words.

Conclusions

An effort will now be made to answer the two questions raised at the beginning of this study:

1. Can children with a mental age of four years learn to read in a manner comparable with the success of average first-grade children? The answer is that at least some children of this mental age can do so. Individual differences, however, are very large. They range from ability to recognize 269 words out of context after about four months' instruction, to only 20

words; from ability to read like an average first-grade child who has completed a full school year's work to almost complete failure, as measured by both standardized and unstandardized primary grade reading tests. With an additional two months' unsupervised practice in reading it is possible for such a child to read as well as an average eight-year-old child in the second grade.

2. Do bright, average, and dull children, all of the same mental age, learn to read equally well under the same experimental conditions? So far as the subjects of this experiment are concerned, this question must be answered in the negative. The bright three-year-old group was very superior to the other two groups; the average four-year-old group was appreciably superior to the dull five-year-old group.

In addition to the two main conclusions above stated, the results of this experiment suggest also the following:

3. An environment that stimulates an interest in books and stories seems to prepare the way for the arousal of a desire to read.

4. It is possible to establish excellent habits in eye-movements while reading in a subject at least as young as three and one-half years of age.

5. Early training in reading has had no obviously harmful effect on the subjects. On the other hand, it has served to increase their interest in books and stories.

6. The early training in reading has favorably affected later progress in reading in the first grade.

7. It may be that the special reading method developed for this experiment, with its great simplification of difficulties and its definite training in the use of

cues, will prove useful in teaching both normal and subnormal children in the public schools.

8. In view of the results of this experiment, it seems that the great amount of time and energy spent on the teaching of reading in the first grade is unnecessary. If a mental age of four years can accomplish what it did with such brief daily lessons in this experiment, it would seem that a mental age of six should accomplish a far greater amount with ease and pleasure. The greater power of voluntary attention should compensate to some extent at least for the larger size of the class. A better appreciation of possible causes of difficulty in perceiving words (until more is definitely known about them), and a systematic effort to help the children to master these difficulties, would no doubt produce better results. It hardly seems necessary to add that ability grouping is of great importance.

9. It is desirable to use standardized reading tests in the first grade as this is the best means of enabling the teacher to form an adequate judgment of the progress of her pupils. Her subjective opinion is frequently in error.

10. The cues most frequently used in the recognition of words in this experiment appear to be the first part of the word, the geometric shape of the word, the last part of the word, oddly shaped letters, and association in context.

11. There is need for further experimentation to test out the indication that left-to-right inversion of words causes very young children little confusion. This will throw much needed light on the question of perception of words themselves, and also on the

broader question of spatial perception in general. It will indeed be interesting to find out if the geometric form of the word is independent of its absolute spatial position.

APPENDIX I

Brief Sketches of Subjects and their Attitude toward Reading

The Bright Group

Jean (CA 3-3; MA 4-6). Jean was a happy, industrious child, generally quiet, but talkative at times. She had a slight physique as the result of a severe attack of infantile paralysis which had left her completely paralyzed in every limb. Thanks to devoted nursing, she had practically recovered complete use of her arms, but the motor coordination of her legs was not yet normal. She had difficulty in climbing stairs, and when she stooped she was apt to topple over. Illness always affected this weak spot. While a subject of this experiment she had several illnesses. She had influenza during the Christmas vacation, a light attack of measles two or three days after she had been given the final tests, and a childish upset during the time of the retests. Reference to this and its effect on her motor coordination will be found in connection with the Pressey Test. Despite this, she was an extremely energetic child, and showed an unusual eagerness in learning to read. Her enthusiasm for reading never abated. She would willingly have had two or more lessons a day. She played no less absorbedly. She had an unusual memory for verse and showed considerable ability in drawing.

Paul (CA 3-4; MA 3-10). Paul was a shy and affectionate child, twin to Peter. He was left-handed. He was active and interested in play, interested in books and stories, and enjoyed having the latter read to him. His interest in reading was somewhat slow in being aroused and proved rather fluctuating in character. When it was good he showed real ability. On a whole, however, it was rather of a passive nature, and at times, he seemed to read to please the experimenter rather than for his own enjoyment. His voluntary attention span was short, and he had to be watched for signs of fatigue.

Peter (CA 3-4; MA 4-0). Peter was Paul's twin. Outsiders found it almost impossible to distinguish between them, but they were not really alike, once one knew them. The parents, one of whom was a university professor and the other a university graduate and student of psychology, stated they had no reason to believe that Paul and Peter were identical twins. In fact, from their own observation, they considered them to be quite different. Both twins

showed considerable initiative. Peter was much more vigorous and independent than his twin, yet, withal, shy. He was right-handed. He, too, was very much interested in books and stories. From the beginning he showed a much keener interest in reading than did Paul, and kept his interest to the end, never showing signs of fatigue.

Kate (CA 3-3; MA 4-2). This subject had a precocious command of language and surprising poise. She was an imperious child, very independent and possessed considerable initiative. She had an over-powering passion for building with bricks. She was also imaginative and frequently had an imaginary playmate with her. She had the weakest voluntary attention span of any subject. Her interest in reading was very slow to develop and she quickly showed signs of fatigue. Because of this, a strict watch was kept, and her first lessons rarely lasted more than four or five minutes. As soon as she showed symptoms of loss of interest she was sent off to play. Later, her interest gradually developed and she did quite well with her work, being able to take a full lesson each day without growing weary. She was extremely fond of stories.

Sally (CA 3-0; MA 4-2). Sally radiated health and happiness. She was the biggest and youngest child in this group. She had a very joyous disposition. She was very vigorous and active and enjoyed stories. Right from the start she showed unusual interest and ability in reading. Her great enthusiasm for reading, like Jean's, never abated. Her reading lesson was always much too short.

The Average Group

Leo (CA 4-2; MA 3-8). Leo was a tall lad. He had a very narrow palate, and suffered from adenoids and enlarged tonsils. He was curiously inert, and almost abnormally quiet. He never displayed signs of initiative. Instead of getting playthings for himself he would sit and watch the other children until the experimenter suggested that he should do this or that. In the earlier lessons he showed considerable interest and ability, but his progress soon slowed up with increasing difficulties. Yet he always volunteered for his lessons as if they gave him pleasure. A variation of method might have overcome his difficulty.

Elsa (CA 4-6; MA 4-6). Elsa was a tall, slight child, timid, rather nervous, and easily upset. She was very observant and a great chatter-box. Her attention, easily distracted, flitted with amazing rapidity from one thing to another. Yet, when called back to the thing in hand, she showed a surprising comprehension. She

showed an early and strong interest in reading and never lost it. She was making very excellent progress when she became ill, and was forced to lose the whole of the last week's instruction. This could not have happened at a more unfortunate time. She was just on the threshold of fluent reading, that is, she was on the point of being able to attack independently a page of unseen material. Her illness put a stop to this. She was given the final tests in her home, but she was still very pale and upset. She was extremely restless and impatient, and her attention was flightier than ever.

Lela (CA 4-4; MA 4-6). Lela was very short for her age, but very sturdily built. She radiated health and happiness. In marked contrast to Elsa, she had a very even disposition, was very vigorous and dependable, and possessed considerable "common sense." She displayed a great deal of initiative. She was close to being the perfectly normal child. At the beginning she displayed a great interest in, and enthusiasm for reading, but this was slightly dulled toward the last and she did not fulfill the expectations held out for her.

Dora (CA 4-1; MA 4-2). Dora was a tall, well-developed child. She was somewhat spoiled and was not at all pleased unless she was the center of attraction. At times very affectionate, at other times she displayed a very violent temper. She became quickly discouraged at the success of several of the other subjects in this group and withdrew from the reading lessons. She was told she did not have to do the work if she did not want to. Occasionally, when a subject passed some remark about her behavior, the experimenter said, "Dora does not have to read if she does not want to, but she does not know what she is missing. It's fun to read, isn't it?" Other than this, no pressure was brought to bear on her. She finally voluntarily asked to have her reading lessons again. The time left, unfortunately, was too short to give her a chance to show what she could do, but she gave evidence, at times, of real interest and fair ability.

The Dull Group

George (CA 5-0; MA 3-10). George was a tall, slight child, of an extremely happy disposition, and not at all self-assertive. He was very anxious to please. He had been somewhat starved at home for playthings. He had never been allowed to use scissors, and did not know how to handle them at first. He had never been told

Mother Goose nor fairy tales. He showed an early and abiding enthusiasm for reading. He would gladly have had more and longer lessons.

Edgar (CA 5-6; MA 4-2). Edgar was a well-developed child, generally rather quiet. He did not begin to talk until he was two and one-half years old, and his speech was still somewhat indistinct. He soon became interested in books and showed quite a keen interest in reading, which was well-sustained until towards the last, when frequent absences due to illness in the home rather spoiled his progress.

Donald (CA 5-1; MA 4-2). Donald was a well-developed child. He was a foster child, having been abandoned as an infant by his parents in the automobile of the man and woman who adopted him. He was a great trial to them because of his wilfulness and disobedience. He showed considerable initiative and constructive ability, so much so that the writer felt that he had better mental ability than the others of this group, and better than his Binet test showed. This proved later to be the case. He, too, was found to have an imaginary play-world. He displayed little interest in reading and made little progress.

Bell (CA 5-0; MA 3-8). Bell was slightly built. She had a silly giggle and talked so incessantly and in such mutilated speech that it sounded like nonsense. She must be considered a case of borderline deficiency. She became eagerly interested in reading and sustained her interest rather steadily throughout. She would willingly have had more frequent lessons, but her attention was not very dependable.

APPENDIX II

Description of Name- and Color-Sorting Games and of Cancellation Games

First name-sorting game. Each subject was given an envelope on the bottom right hand corner of which was printed his own name in Type No. 2. It contained fifteen cards, 3 inches by 1 inch, on ten of which was printed his own name, and on each of the other five a name that was not his. The subject was told to sort the cards into two piles, placing in the one pile all the cards bearing his own name, and in the other all the cards bearing a name other than his own name. His attention was called to his name on the envelope, and he was advised to match the name on the card with this "model" name whenever he was in doubt. The first subject to get his cards correctly sorted was allowed to "read" a book.

Second name-sorting game. The same envelopes were distributed to the subjects, but this time, in addition to the ten cards bearing their own names, there were five cards bearing a name beginning and ending with the same letter as their own names, e.g., Jean and John; Lela and Lena. The purpose was to force the subject who was using only the first or last letters as distinguishing cues to use some other feature in addition. The procedure was exactly the same as in the previous game.

First color-sorting game. Cards 4 inches by 1 inch were cut out. The left half of each card was colored either blue, yellow, red, or green, and on the right hand side was printed the name of the color. Five cards of each color were shuffled and put into an envelope. Each subject was given an envelope and asked to sort the cards into piles; all the blue cards in one pile, all the yellow in another, and so on.

Second color-sorting game. One card of each color was put in an envelope. Five cards with the color names on the right hand side but uncolored on the left hand side were added for each color. Each subject was given a complete set. The four colored cards were used as models. The subject's attention was called to the printed words and its meaning explained: "The word beside the blue color says blue," etc. He was then told to select an uncolored card, match its name to one of the four colored cards, see what color was on this card, and color accordingly with his crayons. This gave practice in matching words.

Third color-sorting game. Twenty of the above colored cards, five of each color, were cut in half, the colored portion being cut off from the name portion. One uncut colored card of each color was laid before each subject as a model. He was instructed to find all the blue cards and the cards bearing the word "blue" and place them below the blue model, etc. This again gave practice in matching words.

Cancellation games. These cancellation "games" were printed on paper by means of Type No. 2. Each subject was provided with a copy. The instructions were: "Find every word that says *can* and draw a line around it."

Card A. Cancel *can*.

boy	can	girl	hop
can	hop	can	hop
girl	this	boy	can
hop	is	can	girl

Card B. Cancel *hop*.

hop	run	boy	make
run	hop	drew	hop
make	red	hop	run
hop	run	make	is

Card C. Cancel *run* in another copy of Card B.

Card D. Cancel *make* in another copy of Card B.

Card E. Cancel *cut*.

cut	has	red	dog
red	*make*	*has*	*cut*
cut	has	cut	red
has	dog	red	cut

Card F. Cancel *out* and *is*.

out	her	girl	is
her	make	her	out
out	her	is	out
is	dog	her	out

Card G. Cancel *her* on another copy of F.

Card H. Cancel *her* and *she* from the following printed on the blackboard. (Each child was permitted to cancel one word of each.)

make	girl	her	she
she	her	has	make
has	he	her	she
draw	she	make	her
her	has	she	draw

Card I. Cancel *girl, dress,* and *dog* (at seats).

dog	dress	girl	horse
girl	horse	dog	dress
horse	dog	girl	dog
dress	girl	doll	dress

Card J. The following questions were printed on the blackboard. The subjects had to underline the correct answer. Each subject read and answered one question.

Can a dog run?	Yes	No
Can a cow hop?	Yes	No
Can a boy hop?	Yes	No
Can a door run?	Yes	No
Can you run?	Yes	No

Card K. The following was carried out as above.

Can a horse run?	Yes	No
Can a table hop?	Yes	No
Can a rabbit run?	Yes	No
Can a boy hop?	Yes	No

Card L. The following was carried out as above.

Can a cow hop?	Yes	No
Can a doll run?	Yes	No
Can a table run?	Yes	No
Can a horse hop?	Yes	No

Card M. The following questions were printed on a sheet of paper. A number of small cards bearing *Yes* and *No* were given to each subject. He had to place the correct answer after each question.

Can a dog run?
Can a horse hop?
Can a girl run?
Can a table hop?

Card N. In the following, done at his seat, each subject had to underline:

dog in the first line
make in the second line
after in the third line
with in the fourth line

run	dog	is	can	girl	
this	make	the	black	has	
she	father	he	after	mother	
to	ball	it	with	play	boy

APPENDIX III

A. Description of Tests Given during the Course of the Experiment

1. *Vocabulary tests.* All vocabulary tests were administered in exactly the same way. All words encountered in the reading lessons were printed on separate cards by means of Type No. 1. The words for each test were entered in a notebook. The word cards were exposed one at a time in the same order, and the subject's responses were recorded. The subjects were taken individually.

2. *Unstandardized reading tests*

Test 40. This is a dog.
He has a rabbit.
She is a little girl.
Cut out the father.

Test 50–I. Billy has a dog.
It is brown.
It can run.
It likes to play with Billy.

Test 50–II. My mother has a cat.
The cat has a kitten.
She is white.
She runs after the ball.

Test 50–III. I have a doll.
The girl wanted to live in my house.
He said, I see my father.

Each subject was given Test 40 when he reached his fortieth, and Test 50 when he reached his fiftieth reading lesson. Each part of Test 50 is made up of increasingly difficult material. These tests were printed on white bristol-board by means of Type No. 2. Each subject was asked to read these selections as well as he could. No help of any kind was given. Time and errors were noted.

B. Description of Final Reading Tests

1. *Improvised test*

Form I. This is a boy.
He has a horse.
It is black.
It can run.

Form II. I like this little girl.
She lives in a pretty house.
She plays with a big doll.
She said, My doll can sleep.

These two tests were printed on white bristol-board by means of Type No. 2. Form I was made up of familiar, and Form II of less familiar words. Each subject was asked to read these selections as well as he could. No help of any kind was given. Time and errors were noted.

2. *Wag and Puff Test*

a. Reading test

Once there was a kitten.
The kitten was yellow.
One day the yellow kitten
saw a dog.
She was afraid.
She was afraid of the dog.
She ran.
She ran up a tree.
The dog ran after the kitten,
but he did not get her.

b. Questions

1. What color was the dog?
2. What color was the kitten?
3. Why did the kitten run?
4. Why did the dog not get the kitten?

The above test will be found on pages 242-244 of the First Grade Manual for the Child's Own Way Series (48).

Directions. (*a*) "I have here a very short story about a kitten and a dog that you have not read before. Read it silently and find out what the kitten and the dog did." (*b*) "Now, please read the story to me." The questions were asked after the story had been read orally since these subjects did not know how to read silently. Unknown words were named for them. Time and errors were noted.

3. *The Merrill Preliminary Tests*

Test 1. The cat saw a bird.
It flew away.
The kitten saw it, too.
They went to the barn.
We gave them milk for dinner.

Test II. The girls saw a nest.
It was in a tree.
Four little birds were in the nest.
Soon they could fly.
One day they flew away.

Test III. The dove saw a white cow.
It said, "See my pretty wings!
Don't you wish you could fly?"
The cow said, "I don't want to fly!
I give milk to boys and girls."

Directions. The subjects were asked to read each of these tests as well as they could. Unknown and wrongly named words were named for them. Errors and time were recorded. The above selections are used preliminary to the Merrill Reading Test proper. They were selected because they contained more or less familiar vocabulary, although some of the words were entirely unfamiliar to the subjects. In addition, they were brief, well printed, each on a separate card.

4. *Detroit Word Recognition Test*

Directions. The subjects were asked to read each word (or phrase), to find the picture which represented it, and to draw a line from the word to the picture. The procedure described in the manual of directions was strictly adhered to.

5. *Pressey First Grade Reading Test*

Directions. The subjects were asked to select and encircle the words and sentences named by the experimenter as described in the direction sheet.

6. *Puff is Sleepy.* (Read by Sally and Jean only.)

a. Reading test:

One day Puff went to sleep.
She went to sleep on a hood of a big black car.
The big black car was Billy's father's.
By and by Puff woke up.
She jumped down to the ground.
With her two little ears,
Puff heard something.
It was Sally.
With her one little nose,
Puff smelled something.
It was milk.

With her four little feet, she ran.
With her two little eyes,
she saw the milk.
With her one little pink tongue,
she lapped up the milk,
lap, lap lap.
By and by she went back to the car.
Then she went to sleep.

b. Recognition of the following words and phrases:

By and by	so	Mother
May I go	too	very fast
a big black car	Father	ran after
Once there was	Then	One day
a little kitten	Wag	He said
a little brown dog	Billy	He saw
in the morning	Sally	He ran
we are going	It is	it was
I am going	Puff	came
He wanted	But	very tired
a little boy	and	something
a little girl	She	shall
began	again	green

Directions. "Here is a story. Can you tell me what it is about? Please read it to me." Record time and errors.

The above test is to be found on pages 209 and 210 of the First Grade Manual of the Child's Own Way Series. The phrases are on page 183 of the same manual. The story is a rearrangement of a story read in the Wag and Puff Primer some time before. It contains several words not in the early story.

7. *The Boy Who Ran Away.* (Read by Sally and Jean only. This selection is found on pages 274 and 275 of the above-mentioned manual).

Once there was a boy. He was a little boy. The little boy liked to run away. One day he ran down the road as fast as he could go.

Soon he came to some woods. It was beautiful in the woods. The trees were green, and the birds sang. But the little boy was not happy. He was a long way from home. He sat down and began to cry. By and by a dog came along. As soon as the boy saw the dog he stopped crying. The dog

began to run out of the woods. The little boy ran after him.

At last they came to a house. The dog stopped running. The little boy stopped running, too. The little boy saw his mother. He ran to her and said, "I am so glad to be home. I shall never run away again." He was very happy. His mother and his dog were happy, too.

Directions. First, the subjects were given three minutes in which to read this story silently. Next, they were asked the following questions about the story: (*a*) What did this little boy like to do? (*b*) Where did he run to? (*c*) Why did he cry? (*d*) How did he find his way home? (*e*) What did he say to his mother when he got home? (*f*) Did this happen in the summer or in the winter? (*g*) How do you know?

Next, they read the selection orally, and time and errors were noted. Next, they again answered the above questions. Finally, they read the following list of words given on page 275 of the aforementioned manual:

he	sat	crying	his
way	fast	out	happy
road	some	by	never
down	home	as	too
could	long	came	last
woods	stopped	her	mother
go	cry	running	after
run	along	him	from

Merrill First Grade Reading Test. (Read by Sally only.)

Once a cat and a fox were in the woods. "I am not afraid of dogs," said the fox. "They can't catch me. I know a great many ways to get away. But how would you get away if a dog should run at you?"

"I know one way to get away," said the cat.

"Only one way to get away!" said the fox. "Why, I know ten ways! I know ten times ten ways. I am sorry for you, and I would tell you some of my ways but you would tell others."

Just then some dogs came running at them. The fox ran this way and that, but the dogs caught him. The cat ran up a tree and the dogs could not get her.

This is the story Sally read while her eye-movements were being photographed. After reading it she was asked the following questions: (*a*) What is the story about? (*b*) What did the fox say? (*c*) How many ways did the cat know of to get away? (*d*) How many ways did the fox know of? (*e*) Who came running then? (*f*) What happened to the fox? (*g*) Did the cat get caught? (*h*) Why not?

APPENDIX IV

Sentences Read by All Subjects during the Pre-Primer Period

1.	This is a girl.	Sentences 1 to 36 were introduced as described in Stage 1.
2.	This is a boy.	
3.	This girl (boy) is (name).	
4.	This is a girl	
5.	This girl is (name).	
6.	This is (name).	
7.	(Name) is a girl (boy).	
8.	This is (name).	Sentences 8 to 11 were printed in small type in a booklet and illustrated by the subjects.
9.	(Name) is a girl (boy).	
10.	This is a girl.	
11.	This is a boy.	
12.	Janet can run.	
13.	Can (name) run?	*S* ran round room.
14.	Yes, (name) can run.	
15.	Can a dog run?	
16.	Yes, a dog can run.	
17.	can a horse run?	
18.	Yes, a horse can run.	
19.	Janet can hop.	
20.	can (name) hop?	*S* hopped round room.
21.	Yes. (name) can hop.	
22.	can a dog hop?	*S* answered orally.
23.	can a horse hop?	
24.	can a cow hop?	
25.	can a rabbit hop?	
26.	can a boy hop?	
27.	Yes, a boy can hop.	
28.	can a girl run?	
29.	Yes, a girl can run.	
30.	run to the cow (name).	*S* ran to the picture of the cow, etc.
31.	run to the horse.	
32.	hop to the dog.	
33.	hop to the chair.	
34.	run to the door.	
35.	Janet can draw.	
36.	can (name) draw?	

37. draw a horse (name). 38. make a horse green. 39. draw a dog. 40. make the dog blue. 41. draw a horse. 42. make the horse yellow.	Sentences 37 to 42 were presented as in Stage 2.
43. draw a rabbit. 44. make the rabbit yellow. 45. cut out the rabbit. 46. Jimmie can draw a cow. 47. Jimmie can draw a rabbit. 48. Jimmie can cut out a horse. 49. cut out a dog (name). 50. cut out a horse.	Sentences 43 to 53 were presented as in Stage 3.
51. draw a chair. 52. make a chair blue. 53. cut out the chair.	Sentences 51 to 53 in small type for seat-work.
54. This is Janet. 55. She is a girl. 56. She is a little girl. 57. This little girl is Janet. 58. This little girl (boy) is (name).	Picture of *S* was introduced.
59. draw a dog. 60. make a dog yellow. 61. cut out the dog.	Small type for seat-work.
62. This little girl is Janet. 63. She has a dress. 64. She has a blue dress.	
65. This is her dress. 66. make her dress blue. 67. cut her dress out.	Small type for seat-work.
68. She has a hat. 69. She has a blue hat.	
70. This is her hat. 71. make her hat blue. 72. cut her hat out.	Small type for seat-work.

73. This little girl is Mary.
74. This is her dress.
75. Her dress is green.
76. is her dress white?
77. No, her dress is green.
78. has she a black hat?
79. No, she has a white hat.
80. Jimmie is a little boy.
81. is (name) a girl?
82. No, (name) is a little boy.

83. Janet has a doll.
84. her doll has a green dress.
85. her doll has a white dress.
86. has her doll a blue dress?
87. Yes, She has a blue dress.

Sentences 83 to 234 were presented on charts as in Stage 5.

88. This is her doll.
89. make her dress blue.
90. make her hat yellow.
91. make her socks blue.
92. make her shoes black.

Sentences 88 to 92 were in small type only for seat-work.

93. is her hat white?
94. No, her hat is black.
95. has she a red hat?
96. Yes. She has a red hat.

97. This is (name).
98. (name) is a girl (boy).
99. This is Janet.
100. She is a girl.
101. She is a little girl.
102. She has a dress.
103. She has a blue dress.
104. This is her blue dress.
105. She has a hat.
106. She has a blue hat.
107. This is her blue hat.
108. Jimmie has a rabbit.
109. He has an Easter rabbit.
110. He has a little rabbit.
111. This is his rabbit.
112. his rabbit is white.
113. his rabbit can run.
114. his rabbit can hop.

Sentences 97 to 107 were in small type in a booklet, illustrated by *S*.

115. This is his rabbit.
116. cut the rabbit out.
117. cut out a cage.

Small type for seat-work.

118. Janet has a little chicken.
119. her chicken is yellow.
120. This is her chicken.
121. Jimmie has a little chicken.
122. his chicken is black.
123. This is his chicken.

124. make the chicken yellow.
125. cut it out.
126. cut out a cage.

Small type for seat-work.

127. Janet has a table.
128. She has a red table.
129. Jimmie has a table.
130. He has a blue table.
131. She has a red chair.
132. He has a blue chair.
133. She has a red hat.
134. He has a blue hat.
135. her dog is black.
136. his dog is white.
137. This is her dog.
138. This is his dog.
139. This is her table.
140. This is his table.
141. This is his rabbit.
142. This is her rabbit.
143. This is Janet's father.
144. He is Mary's father, too.
145. Janet loves her father.
146. her father loves Janet.
147. Mary loves her father.
148. her father loves Mary.
149. This is Janet's mother.
150. She is Mary's mother too.
151. Janet loves her mother.
152. her mother loves Janet.
153. Mary loves her mother.
154. Her mother loves Mary.

155. This is Janet's dog.
156. her father gave the dog to her.
157. the dog's name is Sport.
158. Sport is black.
159. Janet loves Sport.
160. Sport loves Janet.
161. Janet likes to play with Sport.
162. Sport likes to play with Janet.
163. Mary likes to play with Sport.

164. This is Jimmie's dog.
165. his mother gave the dog to him.
166. the dog's name is Rover.
167. Rover is white and black.
168. Jimmie loves Rover.
169. Rover loves Jimmie.
170. Jimmie likes to play with Rover.
171. Rover likes to play with Jimmie.
172. Rover likes to play with Sport.

173. Mary has a ball.
174. She has a red and green ball.
175. her mother gave the ball to her.
176. She likes to play with her ball.
177. Sport likes to play with her ball.
178. Janet likes to play with her ball.
179. Mary rolls the ball to Janet.
180. Janet catches the ball.
181. Janet rolls the ball to Mary.
182. Mary catches the ball.
183. Mary rolls the ball to Sport.
184. Sport catches the ball.
185. Janet said: Catch the ball, Mary.
186. She rolls the ball to Mary.
187. Mary catches it.
188. Mary said: Catch the ball, Janet.
189. She rolls the ball to Janet.
190. Janet catches it.

This ball game was quite a favorite with the children.

191. father likes to play ball with Janet and Mary.
192. Father said: Catch the ball, Mary.
193. Father said: Mary, roll the ball to Janet.

N.B. the introducton of the capital F in father.

194. She rolls it to Janet.
195. Janet catches it.
196. Father said: Janet, roll the ball to Mary.
197. She rolls it to Mary.
198. Mary catches it.
199. Father said: Mary, roll the ball to me.
200. She rolls it to father.

201. Jimmie likes to play ball with Rover.
202. He throws the ball.
203. Rover runs after it to catch it.
204. He brings the ball back to Jimmie.
205. He likes to run after the ball.
206. He likes to bring it back to Jimmie.
207. Jimmie said: Rover, run after the ball.

208. bring it back to me.
209. Jimmie throws the ball.
210. Rover runs after it.
211. He brings it back.
212. Jimmie said to Sport: Sport, run after the ball.
213. bring it back to me.
214. Jimmie throws the ball.
215. Sport runs after it.
216. He brings it back.

217. This is Mary's cat.
218. her mother gave the cat to her.
219. the name of the cat is Fluff.
220. Fluff is a white cat.
221. Mary likes to play with Fluff.
222. Fluff likes to play with the ball.
223. Mary rolls the ball.
224. Fluff runs after the ball to catch it.
225. She likes to run after the ball.
226. She likes to catch it.
227. Run after the ball, Fluff, said Mary.
228. She throws the ball.
229. Fluff runs after it and catches it.
230. She likes to run after it.
231. She catches it with her paws.
232. She rolls it with her paws.
233. She has little paws.
234. her paws are little.

On the completion of the above charts, the Wag and Puff Primer by Marjory Hardie was introduced.

APPENDIX 5

Individual Daily Records of Amount Read

(The first numbers refer to the numbered sentences to be found in Appendix IV. After the introduction of the Primer, the numbers refer to its pages.)

Lesson	Jean	Paul	Peter	Kate	Sally
1	1	1	1	1	1
2	2	2	2	2	2
3	3	3	3	3	3
4	4-7	4-7	4-7	4-6	4-7
5	8-14	12-15	12-15	6-7	8-13
6	15-20	14-16	16-23	30	14-24
7	21-27	8-12	24-28	match names	25-27
8	28-33	17-20	29-34	match names	28-32
9	34-37	21-24	35-40	1	33-37
10	38-42	25-27	41-46	3, 32, 31	37-40
11	43-47	28-30	46-50	12-15	41-44
12	47-50	31-34	48-53	16-18	45-49
13	50-53	34-35	54-61	18-20	50-53
14	54-65	37-41	62-64	21-24	54-67
15	66-72	41-46	65-69	25-30	68-72
16	73-78	46-48	70-75	35-38	73-80
17	79-87	48-50	76-79	39-42	83-92
18	88-96	51-53	80-84	43-46	108-117
19	118-132	54-57	84-87 93-96	46-48	118-136
20	135-154	57-61	108-117	48-50	137-154 97-107
21	155-163 97-107	62-67	118-123 97-107	51-55	155-163
22	164-172	68-72	127-135	54-58 62-64	164-177
23	173-181	88-92	136-142	62-64	178-184
24	182-187	story	story	62-67	185-196
25	story*	73-77	143-154	65-72	197-207
26	188-200	78-82	155-163	73-79	208-216
27	201-205	83-87	164-172	83-87 93-96	story*
28	206-216	93-96	173-181	108-114	217-228
29	217-230	108-117	182-190	115-132	229-234
30	test	test	test	test	test
31	231-234	118-134	191-200	143-148	primer 1-3
32	primer 1, 2	135-148	201-207	149-161	3-5
33	3, 4, 5	149-163	208-220	story*	7-11
34	7, 8, 9	164-175	221-234	story	12-13
35	10-13	176-185	story*	164-172	14-17
36	15-16	story*	primer 1	173-190	14-19
37	17-18	186-194	1-3	191-197	21-23
38	18-19	195-204	2-4	198-201	21-23, 76

*For a change each subject helped to make up a story to read, the experimenter printing it on the blackboard.

INDIVIDUAL DAILY RECORDS OF AMOUNT READ (*continued*)

Lesson	Jean	Paul	Peter	Kate	Sally
39	19-21	205-212	4-5	202-207	story
40	test	test	test	test	test
41	21-23	212-222	7-9	208-216	76-81
42	21-23	223-234	10-11	primer 1-3	81-85
43	76-79	primer 1-2	12-13	4-7	80-85
44	79-82	3-4	13-15	8-9	87-90
45	83-85	5-9	15-17	10-12	90-95
46	81-85	10-11	17-19	12-13	96-99
47	85-89	12-13	19-20	14-15	94-98
48	90-95	14-15	20-22	16-17	25-29
49	95-99	15-16	22	18-19	30-35
50	test	test	test	test	test
51	finals	16	21-23	19	35-42
52		finals	76-77	finals	42-49
53			finals		49-52
54					54-56
55					57-62
56					62-68
57					68-74
58					74, 101-105
59					106-110
60					test
61					111-114 finals

Lesson	Leo	Elsa	Lela	Dora
1	1	1	1	1
2	2	2	2	2
3	3	3	3	3
4	4-7	4-7	12-21	4-11
5	7-11	8-14	22-30	3-5
6	12-23	15-26	30-35	1, 3, 2
7	24-30	30-34	36-40	6, 7 12, 15
8	31-35	35-39	41-45	19-22
9	36-40	39-42	46-50	23-30
10	41-43	43-48	50-53	31-35
11	44-49	49-53	54-61	36-40
12	54-61	54-58	62-67	41-45
13	62-64	62-67	68-72	46-50
14	65-72	68-76	73-79	51-53
15	73-79	77-83	79-83	54-61
16	80-83 93-96	83-87	83-86	62-65
17	83-85	108-123	86-87 93-107	65-67
18	86-89 97-107	127-142	108-117	68-72
19	108-112	97-107	118-123	73-79

INDIVIDUAL DAILY RECORDS OF AMOUNT READ (*continued*)

Lesson	Leo	Elsa	Lela	Dora
20	113-126	143-158	127-142	97-107
21	127-143	159-175	143-154	108-112
22	143-154	176-184	155-172	113-118
23	155-163	185-200	173-181	118-123
24	164-173	201-210	story	127-131
25	story	211-222	182-190	131-136
26	173-181	223-228	191-205	137-142
27	182-190	229-234	205-211	story
28	191-195	story	212-226	story
29	196-201	story	225-234	143-150
30	test	test	test	test
31	201-204	primer 1	story	primer 1
32	204-208	2-3	primer 1	2-3
33	209-212	4-5	2-3	5-7
34	212-215	7-8	1-4	8-9
35	216-220	9	5-7	10
36	story	10-11	5-7	10
37	story	12-13	8-10	11-12
38	221-225	14	10-12	13
39	226-230	15	8-9	15
40	test	test	test	test
41	231-234	16-17	13-14	finals
42	story	17-18	15-16	
43	primer 1	19	16-18	
44	2	20	18-19	
45	3	20	20	
46	4	21-22	21	
47	5-6	22	22-23	
48	7	21-22	23-21	
49	8-9	23	21-23	
50	test	test	test	
51	10	25	25	
52	11-12	30-32	26	
53	12-13	32,76	30-31	
54	13	76-79	32	
55	15	79-80	76-78	
56	16,17	finals	finals	
57	18-19			
58	20			
59	30			
60	finals			
61				

Lesson	George	Edgar	Donald	Bell
1	1	1	1	1
2	2	2	2	2
3	3	3	3	3
4	4-11	4-11	4-11	4-11
5	12-16	12-16	12-16	6-7 12-13
6	14-22	14-22	13-17	12-17
7	23-31	23-28	18-22	18-24
8	32-35	30-36	23-24	25-29

INDIVIDUAL DAILY RECORDS OF AMOUNT READ (*continued*)

Lesson	George	Edgar	Donald	Bell
9	36-40	37-40	30-33	30-37
10	41-44	41-47	31-32	37-40
11	45-50	48-53	30-34	40-45
12	50-53	54-61	35-40	45-50
13	54-61	62-67	41-46	51-53
14	62-67	68-79	49-53	54-58
15	68-72	83-87	54-58	54-58
16	73-80	93-107	62,55	30-32
17	80-83	108-117	drill	54-58
18	83-87	118-123	63-72	63-67
19	93-107	127-142	73-76	97-105
20	108-117	143-148	97-107	68-77
21	118-123	149-163	83-87	83-87
22	127-142	164-178	108-123	93-69 108-123
23	143-156	story	97-107	127-130
24	155-167	179-190	127-134	127-134
25	164-173	191-204	135-142	135-142
26	story	205-217	143-151	143-150
27	173-181	217-223	151-160	150-154
28	182-190	224-234	161-166	155-163
29	191-200	story	167-172	164-170
30	test	test	test	test
31	201-204	217-226	173-184	171-175
32	205-211	227-234	217-226	217-226
33	212-216	primer 1-2	227-231	227-230
34	217-221	3-4	185-192	173-181
35	222-227	7,8	191-200	182-186
36	227-234	8-11	200-207	187-190
37	217-226	12-13	208-216	190-200
38	227-234	14-15	story	story
39	story	16-17	story	story
40	test	test	test	test
41	primer 1-2	18-19	primer 1-3	primer 1-2
42	3	20	4	3-4
43	4-5	21-22	5,8,7	5
44	7,8	22	9	7-8
45	5,9	23	10-11	9-10
46	10-11	30	12-13	11
47	11-12	31	15	12-13
48	13	76	17	14-15
49	14	77-78	18	15
50	test	test	test	test
51	15-16	finals	game	16-17
52	17		game	17
53	18,19		18,19	18
54	20		finals	18,19
55	18,19			20
56	19			30-31
57	20-21			31-32
58	30-31			32-33
59	31			76
60	finals			finals

References*

A. Comparisons of Bright and Dull Children of the Same Mental Level

*1. Baker, H. J. Characteristic differences in bright and dull children. Bloomington, Ill.: Pub. School Publ. Co., 1927. Pp. 118.

2. Chapman, J. C., & Dale, B. A further criterion for the selection of mental test elements. *J. Educ. Psychol.*, 1922, 13, 267-276.

3. Colvin, S. S. Notes on certain aspects of learning. *Psychol. Bull.*, 1915, 12, 67-68.

*4. Cunningham, K. S. The measure of early levels of intelligence. *Teach. Coll. Contrib. Educ.*, 1927, No. 259. Pp. v+74.

*5. Fox, E. J. The diagnostic value of group tests as determined by the qualitative differences between normal and feebleminded children. *J. Appl. Psychol.*, 1927, 11, 127-134.

6. Jones, C. T. Very bright and feeble-minded children. The study of qualitative differences. *Tr. School Bull.*, 1919-1920, 15-16, 137-141; 153-164; 169-180.

7. Merrill, M. A. Achievement in mentally retarded children. *Comp. Psychol. Monog.*, 1924, 2. Pp. 100.

8. Murdoch, K. Rate of improvement of the feeble-minded as shown by standardized educational tests. *J. Appl. Psychol.*, 1918, 2, 243-249.

9. Taylor, G. An inventory of minds of six and seven mental age. *Teach. Coll. Contrib. Educ.*, 1923, No. 134. Pp. 147.

10. Terman, L. M. The intelligence of school children. Boston: Houghton Mifflin, 1919. Pp. xxvii+317.

11. Torgerson, T. L., & Shuman, I. The variability of the acomplishments of pupils of the same mental level. *J. Educ. Res.*, 1925, 11, 132-136.

*12. Wilson, F. T. Some achievements of pupils of the same mental ages but different intelligence quotients. *J. Educ. Res.*, 1926, 14, 45-53.

*An asterisk marks those studies which were not published until after this experiment was begun in February, 1926.

13. WOODROW, H. Practice and transference in normal and feeble-minded children. *J. Educ. Psychol.*, 1917, 8, 85-96; 151-165.

B. THE TEACHING OF READING AND ITS RELATION TO CA, MA, ETC.

14. ANON. An experiment in infant education. (Introduction by L. M. Terman.) *J. Appl. Psychol.*, 1918, 2, 219-228.

15. ARTHUR, G. A quantitative study of the results of grouping first grade classes according to mental age. *J. Educ. Res.*, 1925, 12, 173-185.

16. BOGGS, L. P. How children learn to read. *Ped. Sem.*, 1905, 12, 496-502.

17. BROOKS, F. D. Chronological age as a factor in reading achievement. *School & Soc.*, 1924, 20, 826-828.

*18. ————. Applied psychology of reading. New York: Appleton, 1926. Pp. xvii+278.

19. BROWN, M. W. A study of reading ability in pre-school children. M. A. thesis, Stanford Univ., 1924.

20. COX, C. Genetic studies of genius. Vol. 2. The early mental traits of three hundred geniuses. Stanford University, Calif.: Stanford Univ. Press, 1926. Pp. 842.

21. DEWEY, J. The primary education fetish. *Forum*, 1898, 25, 315-328.

22. DICKINSON, C. E. A study of the relation of reading ability to scholastic achievement. *School Rev.*, 1925, 33, 616-626.

23. DICKSON, V. E. What first grade children can do as related to what is shown by mental tests. *J. Educ. Res.*, 1920, 2, 475-480.

24. ————. Mental tests and the classroom teacher. Yonkers, N. Y.: World Book Co., 1923. Pp. xiv+231.

25. DOUGHERTY, M. L. How to teach phonics. Boston: Houghton Mifflin, 1923.

26. FERNALD, G., & KELLER, H. The effect of kinaesthetic factors in the development of word recognition in the case of non-readers. *J. Educ. Res.*, 1921, 4, 355-377.

27. FLEMING, G. L. Instructional costs in the first six elementary grades. M. A. thesis, Univ. Chicago, 1916.

28. GATES, A. I. An experimental and statistical study of reading and reading tests. *J. Educ. Psychol.*, 1921, 12, 303-314; 378-391; 445-464.

29. ———. Correlations of achievement in school subjects with intelligence tests and other variables. *J. Educ. Res.*, 1922, 13, 129-139; 223-235; 277-285.

30. ———. The psychology of reading and spelling with special reference to disability. *Teach. Coll. Contrib. Educ.*, 1922, No. 129. Pp. v+106.

31. ———. General and specific effects of training in reading. *Teach Coll. Rec.*, 1924, 25, 98-123.

32. ———. Problems in beginning reading. *Teach. Coll. Rec.*, 1924-1925, 26, 572-591.

33. ———. The supplementary device versus the intrinsic method of teaching reading. *Elem. School J.*, 1924-1925, 25, 775-786.

34. ———. Experimental investigations of learning in the case of young children. *J. Educ. Res.*, 1925, 12, 41-49.

35. ———. Functions of flash card exercises in reading. *Teach. Coll. Rec.*, 1925-1926, 27, 311-326.

*36. ———. Method and theories of teaching reading tested by studies of the deaf. *J. Educ. Res.*, 1926, 14, 21-33.

*37. ———. A study of the rôle of visual perception, intelligence and certain associative processes in reading and spelling *J. Educ. Psychol.*, 1926, 17, 433-445.

*38. ———. Studies of phonetic training in beginning reading. *J. Educ. Psychol.*, 1927, 18, 217-226.

*39. ———. The improvement of reading: a program of diagnostic and remedial methods. New York: Macmillan, 1927. Pp. xii+440.

*40. Gates, A. I., Batchelder, M., & Betzner, J. A modern systematic versus an opportunistic method of teaching. *Teach. Coll. Rec.*, 1925-1926, 27, 679-701.

41. Germane, C. E. Silent reading. New York: Peterson, 1922.

*42. Goodenough, F. L. A preliminary report on the effect of nursery school training upon the intelligence test scores of young children. *27th Yrbk. Nat. Soc. Stud. Educ.*, 1928, 361-369.

43. Gray, C. T. Types of reading ability as exhibited through tests and laboratory experiment. *Suppl. Educ. Monog.*, 1917, 1.

44. Gray, W. S. Principles of method in teaching reading as

derived from scientific investigations. *18th Yrbk. Nat. Soc. Stud. Educ.*, 1919, Pt. 2, 25-61.

45. ———. Summary of investigations relating to reading. *Suppl. Educ. Monog.*, 1925, No. 28.

*46. ———. Summary of reading investigations from July 1, 1924, to June 30, 1925. *Elem. School J.*, 1925-1926, 26, 449-459; 574-584; 662-673.

47. HALL, I., & CROSBY, A. A study of the causes of inferior scholarship in low first grade. *J. Educ. Res.*, 1926, 14, 375-383.

48. HARDY, M. First grade manual for the child's own way series. Chicago: Wheeler Publ. Co., 1926.

49. HOLMES, H. W. Time distribution by subjects and grades in representative cities. *14th Yrbk. Nat. Soc. Stud. Educ.*, 1915, 21-27.

50. JUDD, C. H. Reading: its nature and development. Chicago: Univ. Chicago Press, 1918.

51. KLAPPER, P. Teaching children to read. New York: Appleton, 1914. Pp. viii+213.

52. KOFFKA, K. The growth of the mind: an introduction to child psychology. (Trans. from the German by R. M. Ogden.) New York: Harcourt, Brace; London: Kegan Paul, 1924. Pp. xvi+383.

*53. LINCOLN, E. A. The reliability of the Stanford Binet Scale and the constancy of the intelligence quotient. *J. Educ. Psychol.*, 1927, 18, 621-626.

54. MEAD, C. D. The age of walking and talking and its relation to general intelligence. *Ped. Sem.*, 1913, 20, 460-484.

55. MERRILL, M. A. Ability of special class children in the three R's. *Ped. Sem.*, 1918, 25, 88-96.

56. MOORE, A. E. Use of children's initiative in beginning reading. *Teach. Coll. Rec.*, 1916, 17, 330-343.

57. National Society for the Study of Education. Eighteenth Yearbook.

58. ———. Twentieth Yearbook.

*59. ———. Twenty-seventh Yearbook. Bloomington, Ill.; Pub. School Publ. Co., 1928.

60. O'BRIEN, J. A. Silent reading. New York: Macmillan, 1921. Pp. xvii+289.

*61. ———. Reading: its psychology and pedagogy. New York: Century, 1926. Pp. 336.

62. Parker, S. C. How to teach beginning reading. *Elem. School J.*, 1921-1922, 22, 15-30; 104-117; 175-188; 255-268.

63. Patrick, G. T. Should children under ten years learn to read? *Pop. Sci. Mo.*, 1898-1899, 54, 382-392.

64. Pennell, M. E., & Cusack, A. M. How to teach reading. Boston: Houghton Mifflin, 1924.

*65. Peyton, E. M., & Porter, J. P. Old and new ways of teaching primary reading. *J. Appl. Psychol.*, 1926, 10, 264-276.

*66. Reed, M. M. An investigation of practices in first grade admission and promotion. *Teach. Coll. Contrib. Educ.*, 1927, No. 290.

67. Sholty, M. A study of the reading vocabulary of children. *Elem. School J.*, 1911-1912, 12, 272-277.

68. Stern, W. The psychology of early childhood. (3rd ed.) (Trans. by A. Barwell.) New York; Holt, 1924. Pp. 557.

69. Stone, C. T. Silent and oral reading. Boston: Houghton Mifflin, 1922.

70. Taussig, F. W. Principles of economics. (2 vols.) New York: Macmillan, 1911.

71. Terman, L. M. The possibilities and limitations of training. *J. Educ. Res.*, 1924, 10, 335-343.

72. Terman, L. M., *et al.* Genetic studies of genius. Vol. 1. Mental and physical traits of a thousand gifted children. Stanford University, Calif.: Stanford Univ. Press, 1925. Pp. xv+648.

73. Theisen, W. W. Factors affecting results in primary reading. *20th Yrbk. Nat. Soc. Stud. Educ.*, 1921, Pt. 2.

74. ———. Does intelligence tell in first grade reading? *Elem. School J.*, 1921-1922, 22, 530-534.

*75. Thompson, H. An experimental study of beginning reading in deaf-mutes. *Teach. Coll. Contrib. Educ.*, 1927, No. 254. Pp. iv+81.

76. Uhl, W. The materials of reading. New York: Silver Burdett, 1924.

77. Watkins, E. How to teach silent reading to beginners. Philadelphia: Lippincott, 1922.

78. WHEAT, H. G. The teaching of reading. Boston: Ginn, 1923.

*79. WILSON, D. The relationship of the intelligence level to the reading ability of children in the first grade. M.A. thesis, Stanford Univ., 1926.

80. WINCH, W. H. When should a child begin school? (*Educ. Psychol. Monog.*) Baltimore, Md.: Warwick & York, 1911. Pp. 98.

81. ———. Teaching beginners to read in England. *J. Educ. Res. Monog.*, 1925, No. 8.

82. WOOLLEY, H. T. The validity of standards of mental measurement in young children. *School & Soc.*, 1925, 21, 476-482.

*83. YOAKAM, G. A. Reading and study. New York: Macmillan, 1928.

84. ZIRBES, L. Attacking the causes of reading deficiences. *Teach. Coll. Rec.*, 1924-1925, 26, 856-866.

C. EXPERIMENTAL LITERATURE ON THE PERCEPTION OF WORDS

85. ARMAIGNAC, —. *C. r. Soc. fr. ophtal.*, 1892, 10, 361.

86. BOWDEN, J. Learning to read. *Elem. School Teach.*, 1911, 12, 21-33.

87. BUSWELL, G. T. Fundamental reading habits. *Suppl. Educ. Monog.*, 1922, 21. Pp. 150.

88. CATTELL, J. McK. Uber die Zeit der Erkennung und Bennennung von Schriftzeichen, Bildern und Farben. *Wundt's Phil. Stud.*, 1885, 2, 635-649.

89. ———. The inertia of the eye and brain. *Brain*, 1885-1886, 8, 295-312.

90. ———. Ueber die Tragheit der Netzhaut und des Sehcentrums. *Wundt's Phil. Stud.*, 1886, 3, 94-127.

91. ———. The time taken up by cerebral operations. *Mind*, 1886, 11, 220-242; 337-392; 524-528.

92. ———. Psychometrische Untersuchungen. *Wundt's Phil. Stud.*, 1886, 3, 305-335; 452-493.

93. ———. The time it takes to see and name objects. *Mind*, 1886, 11, 63-65.

94. DEARBORN, W. F. The psychology of reading. (*Arch. Phil., Psychol., etc.*, 1906, 4.) New York: Science Press, 1904. Pp. 134.

95. ———. Professor Cattell's studies of perception in reading. *Arch. Psychol.,* 1913-1915, 4, 34-45.

96. DODGE, R. The psychology of reading. *Psychol. Rev.,* 1901, 8, 56-60.

97. ERDMANN, B., & DODGE, R. Psychologische Untersuchungen über das Lesen auf experimentellen Grundlage. Halle: Niemeyer, 1898. Pp. viii+360.

98. FINZI, J. Zur Untersuchungen der Auffassungsfähigkeit und Merkfähigkeit. *Psychol. Arbeit.* (Kraepelin), 1900, 3, 289-384.

99. GATES, A. I. A study of the initial stages in reading by pre-school children. *Teach. Coll. Rec.,* 1923, 24, 469-488.

100. GOLDSCHEIDER, A., & MÜLLER, R. F. Zur Physiologie und Pathologie des Lesens. *Zsch. f. klin. Med.,* 1893, 23, 131-167.

101. HAMILTON, F. M. The perceptual factors in reading. *Arch. Psychol.,* 1906, 1, No. 9.

102. HUEY, E. B. Preliminary experiments in the physiology and psychology of reading. *Amer. J. Psychol.,* 1898, 9, 575-586.

103. ———. On the psychology and physiology of reading. *Amer. J. Psychol.,* 1900, 11, 283-302.

104. ———. On the psychology and physiology of reading. *Amer. J. Psychol.,* 1901, 12, 293-312.

105. ———. The psychology and pedagogy of reading. New York: Macmillan, 1908. Pp. 469.

106. JAVAL, E. Essai sur la physiologie de la lecture. *Ann. d'ocul.,* 1878, 79, 97-117, 240-274; 1878, 80, 135-147; 1879, 81, 61-73, 157-162; 1879, 82, 72-81, 159-170. (See 79, p. 251.)

107. LAMARE, —. Des mouvements des yeux dans la lecture. *C. r. Soc. fr. ophtal.,* 1892, 10, 354-364.

108. LANDOLT, —. Nouvelles recherches sur la physiologie des mouvements des yeux. *Arch. d'ophtal.,* 1891, 11, 385-395.

109. MESSMER, O. Zur Psychologie des Lesens bei Kindern und Erwachsenen. *Arch. f. d. ges. Psychol.,* 1903-1904, 2, 190-298.

110. MILES, W. R., & SHEN, E. Photographic recording of eye movements in the reading of Chinese in vertical and horizontal axes. *J. Exper. Psychol.,* 1925, 8, 344-362.

111. MEEK, L. H. A study of learning and retention in young children. *Teach. Coll. Contrib. Educ.*, 1925, No. 164. Pp. ix+96.

112. PILLSBURY, W. B. A study in apperception. *Amer. J. Psychol.*, 1897, 8, 315-393.

113. QUANTZ, J. O. Problems in the psychology of reading. *Psychol. Rev., Monog, Suppl.*, 1897, 2, No. 1. Pp. 51.

114. SCHMIDT, W. A. An experimental study in the psychology of reading. *Suppl. Educ. Monog.*, 1917.

*115. TINKER, M. A. An experimental study of the legibility, perception and eye movements in the reading of formulae. Ph.D. thesis, Stanford Univ., 1927.

116. ZEITLER, J. Tachistoskopische Untersuchungen über das Lesen. *Wundt's Phil. Stud.*, 1900, 16, 380-463.

UNE ÉTUDE EXPÉRIMENTALE DES ENFANTS TRÈS INTELLIGENTS, DES ENFANTS D'INTELLIGENCE MOYENNE, ET DES ENFANTS BORNÉS AU NIVEAU MENTAL DE QUATRE ANS

(Résumé)

On a fait cette expérience dans le but de trouver (1) à quel degré les enfants d'un âge mental de quatre ans pourraient apprendre à lire, et (2) si les enfants très intelligents, ceux d'intelligence moyenne, et les enfants bornés, tous de cet âge mental, apprendraient à lire également bien dans les mêmes conditions expérimentales. On a employé trois groupes d'enfants de l'âge pré-scolaire; un groupe composé d'enfants très intelligents âgés de trois ans, un groupe d'enfants d'intelligence moyenne âgés de quatre ans, et un groupe d'enfants bornés âgés de cinq ans. Il y a eu cinq enfants dans le groupe des intelligents, et quatre dans chacun des autres groupes, celui des enfants d'intelligence moyenne et celui des bornés. Les groupes ont eu les quotients moyens d'intelligence suivants: les intelligents, 128; ceux d'intelligence moyenne, 98,50; les bornés, 77.

Chaque groupe s'est réuni comme groupe, à des heures différentes, cinq jours par semaine, pendant à peu près quatre mois et demi. Chaque sujet a reçu une leçon quotidienne de lecture de dix minutes, et le groupe, comme ensemble, a reçu généralement aussi un court jeu collectif quotidien.

Après une analyse soigneuse des difficultés possibles que l'apprentissage de la lecture pourrait présenter aux enfants de ce niveau mental, on a développé une méthode spéciale de lecture contenant des traits nouveaux. Le trait le plus original a été une série de ce qu'on a appelé des "Formes Massées," lesquelles ont servi à remplir la lacune entre nulle science des symboles imprimés des mots et la science et l'intérêt pour ceux-ci.

A des intervalles réguliers pendant l'expérience, et aussi à sa fin et encore quelques semaines plus tard, on a fait subir des tests complets de vocabulaire et des tests de lecture standardisés et non standardisés pour les classes élémentaires. Les résultats ont montré la grande supériorité du groupe des enfants intelligents de trois ans sur les deux autres groupes, tandis que le groupe des enfants de quatre ans, d'intelligence moyenne, a été un peu supérieur au groupe des bornés. A la fin de l'expérience les enfants intelligents de trois ans ont reconnu une moyenne de 129,40 mots dans le contexte, les enfants d'intelligence moyenne, âgés de quatre ans, 55,30 mots, et les bornés de cinq ans 40,00 mots. Il y a eu de grandes différences individuelles dans chaque groupe. A l'époque des deuxièmes épreuves, le sujet qui avait le plus réussi a su lire, à l'âge de trois ans et sept mois, aussi bien qu'un enfant normal de huit ans. A cette époque on a photographié les mouvements de ses yeux pendant qu'elle lisait une histoire pas du tout connue et on les a trouvés supérieurs à ceux des enfants de la Third Grade.

Pour jeter une lumière très nécessaire sur la question importante de comment les jeunes enfants perçoivent les mots, on a analysé les erreurs faites dans les tests de vocabulaire de cette expérience. Peut-être que le résultat le plus intéressant de cette analyse a été la découverte d'une indication que l'inversion des mots de la gauche à la droite cause peu de confusion chez les très jeunes enfants.

Enfin, on a analysé tous les facteurs du développement et du milieu qui pourraient influer sur une lecture réussie.

DAVIDSON.

EINE EXPERIMENTELLE UNTERSUCHUNG AN KLUGEN, MITTELMÄSSIGEN, UND GEISTIG ZURÜCKGEBLIEBENEN ('DULL') KINDERN AUF DEM NIVEAU DER VIERJÄHRIGEN GEISTIGEN ENTWICKLUNG

(Referat)

Es war der Zweck dieser Untersuchung, festzustellen, 1) bis zu welchem Grade Kinder mit vierjähriger geistiger Entwicklung lesen lernen konnten, und 2) ob kluge, mittelmässige, und geistig zurückgebliebene Kinder, alle in dem selben geistigen Alter, ('mental age'), unter gleichen experimentellen Bedingungen gleich gut lesen lernen würden. Man arbeitete mit drei Gruppen vorschulpflichtiger Kinder, -eine kluge (begabte) drei-Jahr-alte Gruppe, eine normale vier-Jahr-alte, und eine geistig zurückgebliebene ('dull') fünf-Jahr-alte Gruppe. In der 'klugen' Gruppe befanden sich 5, in der 'mittelmässigen' und 'zurückgebliebenen' je 4 Kinder. Die mittleren Intelligenzquotienten der klugen, mittelmässigen, und zurückgeblieben Gruppen betrugen respektiv 128, 98.5, und 77.

Die Gruppen kamen als solche, zu verschiedenen Stunden, 5 Tage in der Woche, während eines Zeitraumes von fast 4 1/2 Monaten zusammen. Jede Vp. erhielt täglich einen Leseunterricht der 10 Minuten dauerte, und die Gruppe als ganze hatte gewöhnlich noch dazu an einem kurzen täglichen Gruppenspiel Teil.

Nach sorgfältiger Analyse der Schwierigkeiten, welche das Lesen lernen Kindern dieses geistigen Niveaus darbieten könnte, wurde eine spezielle, einige neue Besonderheiten enthaltendende Lesemethode entwickelt. Der originellste Zug derselben war eine Serie von sogenannten Massenformen ('mass forms') welche dazu dienten, eine Brücke zu schlagen über die Lücke zwischen völligem Mangel an Kenntnis der gedruckten Wortsymbole einerseits und Kenntnis und Interesse für dieselben anderseits.

Regelmässig, von Zeit zu Zeit, das ganze Experiment hindurch, wie auch am Schlusse desselben, und wieder einige Wochen später, wurden die Vpn. vollkommenen Wortschatzprüfungen und sowohl normierten ('standardized') wie unnormierten Leseprüfungen für die untersten Schulklassen unterworfen. Die Resultate zeigten die grosse Überlegenheit der klugen drei-Jahr-altrigen Gruppe den übrigen zwei Gruppen gegenüber, während die mittelmässige Gruppe der zurückgebliebenen Gruppe etwas überlegen war. Am Ende der Untersuchung erkannte die kluge drei-Jahr-altrige Gruppe durchschnittlich 129.4, die vier-Jahr-altrige Durchschnitts-Gruppe 53.4, und die zurückgebliebene 5-Jahr-altrige Gruppe 40.0 Wörter ohne Hilfe des Kontextes wieder. Es zeigten sich innerhalb jeder Gruppe starke individuelle Unterschiede. Zur Zeit der Wiederholungen der prüfungen, als es drei Jahre und 7 Monate alt war, las die Vp. (ein Mädchen), die den besten Erfolg erzielte, so gut wie ein gewöhnliches Kind von 8 Jahren. Zu dieser Zeit photographierte man ihre Augenbewegungen, während sie eine ihr völlig fremde Geschichte las, und fand diese Bewegungen denen von Kindern in der dritten Schulklasse [an Angemessenheit] überlegen.

Um dringend nötiges Licht auf die wichtige Frage, wie kleine Kinder Wörter wahrnehmen, zu werfen, analysierte man die Fehler, die in den Wörterschatzprüfungen in dieser Untersuchung gemacht wurden. Das interessanteste Resultat dieser Analyse war vielleicht der Befund, dass die Inversion der Wörter von links nach rechts bei sehr jungen Kindern wenig Verwirrung verursacht.

Zum Schlusse wurden alle Entwicklungs- und Umgebungsfaktoren die vielleicht den Erfolg bei dem Lesen beeinflüssen könnten analysiert und besprochen.

DAVIDSON.

$7.00 per volume
Single numbers $2.00

MONTHLY
Two volumes per year

May, 1931
Voume IX, No. 5

GENETIC PSYCHOLOGY MONOGRAPHS

Child Behavior, Animal Behavior, and Comparative Psychology

AN HISTORICAL, CRITICAL, AND EXPERIMENTAL STUDY OF THE SEASHORE-KWALWASSER TEST BATTERY*

From Stanford University

By

PAUL R. FARNSWORTH

*Accepted for publication by Carl Murchison of the Editorial Board and received in the Editorial Office, September 24, 1930.

Worcester, Massachusetts

Entered as second-class matter December 1, 1925, at the post-office at Worcester, Mass., under Act of March 3, 1879

PREFACE

When the writer began this statistical study of the Seashore-Kwalwasser battery he believed that little had been done along similar lines. As the work progressed, a considerable body of data was come upon. It was similar in certain respects to what he had been assembling, but was scattered in numerous articles. Much work was done apparently at about the time his data were being collected.

Errors are likely to creep into a study of this type. Sins of omission are also probable, although many omissions are intentional, i.e., references to the majority of "popular" articles which deal with this battery. There may be readers who will dislike the methods of presenting data. Certain of the experimenters whose studies are herein discussed may conceivably object to the manner in which their materials are treated. The writer defends himself by merely stating that he desired to make comparisons, and arbitrarily chose certain procedures. For his errors and idiosyncrasies he begs charity.

Dr. Robert H. Seashore of the University of Oregon kindly supplied the Iowa pitch and intensity data. Dean Jack Dunlap of the Territorial Normal and Training School, Hawaii, cooperated in obtaining the Kwalwasser scores on Hawaiian subjects. The Hong Kong data were collected by Mr. Charles Voegelin. Professor Norval Church of Columbia collected certain data from the fifth and eighth grades of San José, California. The section of the study which deals with

oriental subjects was financed by a Stanford University grant, a portion of which came from the Laura Spelman Rockefeller Memorial Foundation. Grateful acknowledgment is due Miss E. A. McAnulty of the Los Angeles City Schools and especially Mr. Quinn McNemar of Stanford for statistical advice. Possible statistical errors, however, should be charged solely to the writer. It is obvious that this study would have been impossible but for the cooperation of the subjects, to whom thanks are due. The writer is also greatly indebted to his wife, Dr. H. C. Farnsworth.

PAUL R. FARNSWORTH

STANFORD UNIVERSITY
CALIFORNIA

CONTENTS

INTRODUCTION

Carefully controlled stimuli are obviously needed in any study of human behavior. The Seashore Columbia battery (72) and the Kwalwasser Victor record (39) possibly present stimuli of such a character. Although these sound stimuli show slight physical variations from one rendition of the records to the next, they are at least constant enough to justify careful analysis. The following study was undertaken, therefore, to gauge their utility as psychological tools.

DESCRIPTION OF RECORDS

The Seashore Measures of Musical Talent form a battery of six double-face records. The Sense of Pitch Test attempts to demonstrate the minutest difference in pitch one can detect. One hundred pairs of stimuli are presented. The subject records whether the second member of a pair is higher or lower in pitch than the first member. The test begins with differences (30 d. v.) which are apparent to the majority of subjects. There are ten pairs in this grade of difficulty. There next appear ten which are slightly more difficult to discriminate (23 d.v.). These are followed by eight more groups of ten each which are still more difficult. The magnitudes of difficulty of these remaining eight groups are 17 d.v., 12, 8, ½, 1, 2, 3, and 5. When in doubt as to a judgment, the subject is urged to guess.

The basic theory underlying the Sense of Intensity Test is similar to that attached to the Sense of Pitch Test. The subject listens to sequences of two tones

each, and decides whether the second tone is weaker or stronger in intensity than the first. In both faces of the test the arrangement of times is from easy to difficult. As in the test previously mentioned, one hundred pairs of stimuli are presented.

A third, somewhat similar, test is entitled Sense of Time. This test attempts to disclose the minutest differences in time one can detect. The subject hears one hundred sets of stimuli. Each set is composed of three clicks which mark off (i.e., bound) two intervals of time. The second interval is to be rated as longer or shorter than the first. As in the previously mentioned tests, the sets of stimuli are grouped into tens which are supposed to be of equal difficulty. The magnitude of the temporal differences between the intervals in the first and second groups is 20 hundredths of a second; in the third and fourth groups, 14 hundredths; in the fifth and tenth, 9 hundredths; in the sixth and seventh, 2 hundredths; and in the eighth and ninth, 5 hundredths.

In the Sense of Consonance Test, 50 pairs of simultaneous dyads are offered. The subject compares the consonance of the members of each pair, and rates the second as "better" or "worse" than the first. The specific manner in which "consonance" is defined will be discussed later.

The Sense of Rhythm Test is the newest member of the battery. It is made up of 50 rhythmic patterns which are each played twice, either in the same or in altered form. The subject classifies the repetition as "same" or "different."

"Tonal memory" is tested by the presentation of a series of tones played twice. In the second playing, one note is changed. The subject is to count mentally and record, by number, which note is changed. There are supposed to be five degrees of difficulty—two notes each (played twice) in the first series of ten, three in the next, then four, then five, and, lastly, six. The test thus contains 50 series in all.

The Kwalwasser battery includes tests of "melodic and harmonic sensitivity." Each test consists of 35 items which are to be rated as "good" or "bad." The Melodic Sensitivity Test presents a series of two-measure melodic progressions. The Harmonic Sensitivity Test employs progressions which consist of three chords each.

Reliabilities

Data on Adults

Two types of reliability measures have been employed for these tests—retest and split-half correlations. In this section of the monograph the varieties will be discussed for each test in turn. The correlation coefficients are product-moment *r*'s unless specifically mentioned as different. Probable errors have been figured from the Toops Chart wherever the author in question has not reported them.

Pitch. In the fall of 1919, Gaw (22) gave five members of the Seashore battery to a group of music students. Retests were taken in the spring of 1920. A similar procedure was undertaken for a group of normal-school students. Pitch was found to have reliabili-

ties of .90±.02 for the former group and of .73±.05 for the latter.

With retests on college students Lanier (43) found a reliability of .68±.04. Approximately five months elapsed between tests.

Peterson (60) has been gathering information on the subject of racial differences for some years. In one experiment his pitch retest reliabilities were .88±.02 for white college students (Peabody College) and .77±.02 for the colored (Fisk). In another study he and Lanier (62) found values of .69±.03 for white college students and .58±.04 for colored students.

Junior- and senior-high-school students served as subjects for Brown (7). There was an interval of about four months between tests. A reliability of .71±.03 was reported. In a strict sense high-school subjects are not to be classified as adults on these tests. Their scores would fall between the adult and eighth-grade norms. This study checked a previous one by Ruch and Stoddard (64), who recorded in their book, *Tests and Measurements in High School Instruction,* a reliability of .70±.04.

As reported by McGinnis (51), McCarthy found a retest reliability of .70. The test was given to school children, junior-high, and college students.[1]

Highsmith (31) has studied the records of 59 girls who entered the School of Music of the North Caro-

[1]McCarthy (50*a*) found pitch and memory retests similar to those of Brown. Intensity retests ranged from .48 to .69 (N=58). Grifitts (25*a*) retested 32 subjects after a week with the following coefficients: pitch, .73; intensity, .76; time, .81; consonance, .60; and memory, .83.

lina College for Women in September, 1922. Pitch was found to have a retest reliability of .76. For some reason not stated the record was not given in the normal manner. The sequence was: face A, face B, A, B, and then face B again.

The act of splitting tests into equivalent halves may be accomplished in a number of ways. Lanier (43) has correlated the odd items of the pitch test against the even with a resulting value of .81±.02. The same method, as employed by the present writer, yielded .74±.02 and .75±.02 for two groups from the State University of Iowa, and .76±.02 for a similar Stanford group. The subjects for the three groups were all from large classes in elementary psychology.[2]

Murdock (54) has studied the abilities in the pitch test as found among various racial groups of Hawaii. A reliability of .63 was reported. Its significance is not entirely clear, since the groups were selected by excluding those who scored below a certain arbitrary score.[3]

The Seashore tests are so constituted that the top five rows of tabulated responses should be equivalent

---

[2]Reported in part at the 1927 meeting of the Western Psychological Association.

[3]McCarthy (50*a*), with a population of 58 adults, found values of .77 for pitch, .62 for intensity, and .74 for memory. A vocational class of 35 gave corresponding values of .84, .46, and .80. The consonance value was .48 and the time, .57.

Stanton and Koerth (81*a*) reported odd-even values for two groups (N=157 and 200). The values were: pitch, .60 and .51; intensity, .82 and .78; time, .67 and .66; consonance, .46 and .58; memory, .90 and .85. Retest values on 57 subjects (the retests occurred after three years of intensive musical training and education) were reported as: pitch, .54; intensity, .80; time, .45; consonance, .62; and memory, .83.

to the bottom five. Lanier found that these two portions of the test intercorrelated to the extent of .84$\pm$.02. With the same three groups mentioned above, the writer found values of .71$\pm$.02, .74$\pm$.02, and .73$\pm$.02.

In contrasting the retest and the split-half methods of obtaining reliabilities, it can be said that the latter is probably the more meaningful for at least the following reasons. First, it is a difficult matter to keep up the interest in a second rendition of what to most subjects are uninteresting and fatiguing tests. Second, by the time of the retest the more intelligent subjects may begin to note various cues which will raise their scores. Third, and probably of small importance when the tests are not immediately presented for the second time, is the possible memory carryover.

From the Stanford data, the sigma of the mean was found to be 7.5.[4] If the reliability is assumed to be .75, the *P.E.* of the measurement becomes about 2.5. If the problem arises of predicting the range of any true score, one usually marks off 4 $P.E._M$ on either side of the raw score in question. If this is done, it will be seen that for the range of a true score to be in the highest third of the distribution, a subject must score 94% (100 percentile) or better. To score in the lowest third, he must test 67% (11 percentile) or worse.

While consistent reliabilities of .90 and over are always desirable, the reliabilities recorded here are

[4]The present writer's sigmas are a little smaller, in the main, than those reported by R. C. Larson (47).

fairly good, in the main. The Sense of Pitch Test, then, is a fairly reliable test.

Intensity.[5] The retest reliabilities, as reported by the previously mentioned pitch experimenters, were as follows: Gaw, .55±.06, and .94±.01; Lanier, .60 ±.04; Peterson, no data from his first groups, .75± .03 for white students, and .65±.03 for Negroes; Brown, .65±.04; Ruch and Stoddard, .66±.04; McCarthy, .68; Highsmith, .50±.07 (sequence of record faces was A, B, A).

Lanier found an odd-even reliability for the intensity test of .67±.04. The writer's data on four Iowa groups show values of .72±.02, .61±.03, .65±.02, and .63±.03, while for a Stanford group the figure was .66 ±.03. Lanier's corresponding top-bottom reliability was .61±.04. For the writer, the coefficients were .69 ±.03, .57±.03, .56±.02, .58±.03, and .62±.03, respectively. Two additional groups gave values of .51 ±.03 and .66±.03. The intensity test is so constructed that the phonograph record faces are practically equivalent. Thus a correlation between the two faces should give a measure of reliability. This value was figured for two of the writer's Stanford groups and was found to be .60±.03 and .51±.03.

The sigma of the mean of the Stanford data was found to be 5.6. An assumed reliability of .66 will give a $P.E._M$ of about 2.2. By methods discussed under pitch discrimination it was found that for a subject's true score to fall into the upper third of the distribution, it must be perfect (100 percentile). A score

[5]See footnotes under *Pitch*.

of 77 (8 percentile) or worse will put it in the lowest third.

From a consideration of the reliabilities reported for the Sense of Intensity Test, we must conclude that it has not the reliability status of the Sense of Pitch record.

Time.[6] The retest reliabilities, as reported by the previously mentioned pitch experimenters, were as follows: Gaw, .46±.07, and .56±.07; Lanier, .50±.05; Peterson, no data for his first groups, .45±.05 for white students and .62±.03 for Negroes; Brown, .48±.05; Ruch and Stoddard, .53±.05; McCarthy, no data; Highsmith, .52±.07 (sequence of record faces was A, B, A).

Lanier found an odd-even reliability for the time test of .54±.05. The writer's data on his Stanford group show a value of .51±.04. Lanier's corresponding top-bottom reliability was .59±.04. For the writer the coefficients were .41±.05 and .46±.04. The time test is so constructed that half of the record which includes tabulated Columns A, C, E, G, and I should be comparable to the half made up of Columns B, D, F, H, and J. Correlations between these two halves (by the writer) gave values of .45±.05, and .52±.04.

The sigma of the mean of the Stanford data was found to be 6.9. An assumed reliability of .51 will give a $P.E._{M}$ of about 3.3. By methods discussed under pitch discrimination, it was found that for a subject's true score to fall into the upper third of the distribution, it must be 94 (100 percentile) or better.

[6]See footnotes under *Pitch*.

A score of 62 (3 percentile) or worse will put it in the lowest third.

The above-mentioned reliability coefficients for the Sense of Time Test indicate that, while the test may be usable in certain connections, caution must always be observed when important conclusions are based on a measure which possesses such a low reliability.

Consonance.[7] The retest reliabilities as reported by the previously mentioned experimenters were as follows: Gaw, .49±.08, and .57±.07; Lanier, .54±.05; Peterson, .68±.04 for the white students and .52±.03 for the Negroes of his first groups, and .59±.04 for the white students and .41±.04 for the Negroes of his second groups; Brown, .43±.05; Ruch and Stoddard, .35±.06; McCarthy, no data; Highsmith, .53±.06 (sequence of record faces was A, B, A, B). D. L. Larson (46) reported an unpublished State University of Iowa study on normal-school students in which the reliability was only .36±.06. In a special study of the consonance test, Heinlein (27) tested two groups of students and retested them after a rest period. The present writer has computed rank correlations for Heinlein's groups, and has found them to be: rho=.60, and .46. These values would approximate *r*'s of .62±.07 and .48±.09.

In the studies reported so far, the test directions have been followed with varying degrees of faithfulness. D. L. Larson (46) at the State University of Iowa laboratory has carried out perhaps the most careful retests of the consonance test on record. She especially

[7]See footnotes under *Pitch*.

attempted to make the directions perfectly clear. Retests of two groups, after several months, correlated .63±.02 and .65±.02 with the original scores.

Lanier found an odd-even reliability for the consonance test of .33±.06. The writer's data on his Stanford group show a value of .36±.05. Lanier's corresponding top-bottom reliability was .45±.05. For the writer the coefficient was .33±.05. Between the two faces of the consonance record the latter found a correlation of .16±.05. The consonance record was treated as a preference test in the collection of the Stanford data.

The sigma of the mean of the Stanford data was found to be 3.3 (twice this value if figured in terms of the percentage correct). An assumed reliability of .65 will give a $P.E._M$ of about 1.4. By methods discussed under pitch discrimination, it was found that for a subject's true score to fall into the upper third of the distribution, it must be 82% (99 percentile) or better. A score of 55% (4 percentile) or worse will put it in the lowest third.

The above data indicate that the Sense of Consonance Test is not reliable. While certain experimenters might think it usable, the present writer would hesitate to put much trust in a test which is so unreliable.

Rhythm. The retest reliabilities, as reported by the previously mentioned experimenters, were as follows: Gaw, no data; Lanier, .43±.05; Peterson, no data for his first groups, .50±.05 for white students and .45±.04 for Negroes; Brown, .29±.06; Ruch and Stoddard,

.50±.05; McCarthy, no data. In an unpublished study from the University of Oregon, E. Poorman found retest reliabilities of .53±.07, and .53±.06. The tests were given two weeks apart.

Lanier found an odd-even reliability for the rhythm test of .35±.06. The writer's data on his Stanford group show a value of .47±.04. Lanier's corresponding top-bottom reliability was .28±.06. For the writer the coefficient was .43±.04. The two faces of the rhythm record were found by the latter to correlate at .39±.05.

The sigma of the mean of the Stanford data was found to be 4.1 (twice this value of the percentage correct is considered). An assumed reliability of .47 will give a $P.E._M$ of about 2.0. By methods discussed under pitch discrimination, it was found that for a subject's true score to fall into the upper third of the distribution, it must be perfect. A score of 60% (on Stanford norms—no equivalent Seashore percentile is given, since it will be shown that the Seashore rhythm norms are incorrect) or worse will put it in the lowest third.

The Sense of Rhythm Test is apparently in the class of the Sense of Consonance Test. It is not very reliable.

Tonal Memory.[8] The retest reliabilities, as reported by the previously mentioned experimenters, were as follows: Gaw, .90±.02 and .88±.04; Lanier, .67±.04; Peterson, no data for the first group, .67±.03 for white students and .80±.02 for Negroes; Brown, .59

---

[8]See footnotes under *Pitch*.

±.04; Ruch and Stoddard, .66±.04; McCarthy, no data; Highsmith, .83±.03 (sequence of record faces was A, B, A, B).

Lanier found an odd-even reliability for the memory test of .74±.03. The writer's data on his Stanford group show a value of .83±.02. Lanier's *corresponding top-bottom reliability was* .60±.04. For the writer the coefficient was .81±.02.

The sigma of the mean of the Stanford data was found to be 7.2 (twice this value if the percentage correct is considered). An assumed reliability of .83 will give a $P.E._{M}$ of about 1.9. By methods discussed under pitch discrimination, it was found that for a subject's true score to fall into the upper third of the distribution, it must be 94% (95 percentile) or better. A score of 51% (10 percentile) or worse will put it in the lowest third.

The Tonal Memory Test[9] ranks with Sense of Pitch as a fairly reliable test.

The Kwalwasser Tests.[10] As far as the writer is aware, his is the only study on record of the Kwalwasser reliabilities. He has given retests to a Stanford group after three months with resulting coefficients of .57±.05 for melody and .50±.06 for harmony. In a study of two other groups, he found odd-

[9]The Tonal Memory Test is frequently termed "memory" test in this monograph.

[10]The tests of Melodic and Harmonic Sensitivity are frequently termed the "melody" and "harmony" tests in this monograph. Care should be taken to distinguish these records from others recently issued by Kwalwasser in cooperation with Professor P. W. Dykema of Columbia. The new ones, which are not discussed in this monograph, are Numbers 302, 303, 304, 305, and 306, A and B (Victor).

even reliabilities for the melody test of $.26 \pm .05$ and $.42 \pm .05$. Those for harmony were $.17 \pm .05$ and $.21 \pm .06$. Correlations between the first and second halves of the tests gave values for one group of subjects of $.12 \pm .05$ and $.30 \pm .05$ for melody and harmony, respectively. Needless to say, these two Kwalwasser tests have rather poor reliabilities.

The sigma of the mean of the Stanford melody data was found to be 3.2. An assumed reliability of .42 will give a $P.E._{M}$ of about 1.6. By methods discussed under pitch discrimination, it was found that for a true score to fall into the upper third of the distribution it must be 33 (raw score) or better. A score of 19 or worse will put it in the lowest third.

The sigma of the mean of the Stanford harmony data was found to be 2.7. An assumed reliability of .21 will give a $P.E._{M}$ of about 1.7. It was found that for a true score to fall into the upper third of the distribution it must be 33 (raw score) or better. A score of 17 or worse will put it in the lowest third.

Data on Younger Subjects

McGinnis' Data. McGinnis (51) has given the pitch, intensity, and consonance tests to preschool children. A new technique for giving and scoring the tests was instituted. Retest reliabilities were found to be: pitch, $-.02 \pm .17$; intensity, $.43 \pm .14$; and consonance, $.37 \pm .15$ (rank coefficients). Although the reliabilities are extremely low, it may prove possible to alter techniques, and raise them in the future.

R. C. Larson's Data. By the method of retesting

immediately after the first rendition of each test (see page 317), Mrs. Larson (47) has computed values for grade-school students. There were two groups of eighth-graders with populations of 200 and 100, respectively, who gave correlations as follows: pitch, .83 ±.01, .85±.02; intensity, .80±.02, .83±.02; time, .59 ±.03, .62±.04; consonance, .55±.03, .54±.05; memory, .87±.01, .88±.02; rhythm, .56±.03, .59± .05. The seventh-grade subjects gave these values: pitch, .90±.01; intensity, .78±.03; time, .69±.03; consonance, .57±.04; tonal memory, .90±.01; rhythm, .54±.05. For the sixth grade, the values were: pitch, .82±.01; intensity, .83±.02; time, .80±.02; consonance, .57±.04; memory, .84±.02; rhythm, .42±.05. The fifth-graders gave values as follows: pitch, .84± .02; intensity, .81±.02; time, .85±.02; consonance, .36 ±.05; memory, .84±.02; rhythm, .46±.04. When the sixth-graders were retested, in five months, the retest correlations, with the exception of pitch, were found to be lower than those computed for the same sitting.

Church's Data. Under the writer's guidance, Norval Church has given the Seashore-Kwalwasser battery to fifth- and eighth-grade subjects at San José, California. For the eighth grade the odd-even reliabilities were: pitch, .80±.02; intensity, .75±.02; time, .62±.03; memory, .78±.02; consonance, .43±.04; rhythm, .30±.04; melody, .37±.04; harmony, .21± .05. The fifth-graders gave values of: pitch, .86± .01; intensity, .75±.02; time, .58±.03; memory, .74 ±.02; consonance, .39±.05; rhythm, .45±.04; melody, .40±.04; harmony, .36±.04. These data indicate that

ula, and other values were derived. Second, rows two and three were correlated. This value was averaged with the initial coefficient to make up the "mean of the two correlations." This new value was, in turn, substituted in the formula. The third check was the "best fit" method employed by Holzinger and Clayton. Lanier concluded from his study of this portion of his work:

> "For the majority of tests used here (four of the six) the actual reliabilities given by correlating increased lengths *within one test* lie within one probable error of the corresponding predicated reliabilities. The best fit values are slightly more accurate than this and are so for all six tests."

Lanier also compared the actual and predicted reliabilities of the tests when increased in length at different sittings. Row one was correlated with row one of the retest. Rows one and two were correlated with rows one and two of the retest (sums). This procedure went on until the entire test (all ten rows) was correlated with its retest. As before, three checks were employed. First, the initial r (row one versus row one of the retest) was substituted in the Spearman-Brown formula. Second, a mean was obtained by correlating each row with the corresponding one in the retest (one with one, two with two, etc.), and was substituted in the formula. Third, a "best fit" value was determined.

Concerning his results, Lanier has declared:

> "*a*. In all six tests there is marked over-prediction when either the initial or the mean coefficient is used. *b*. The 'best' value gives predictions that generally include the actual reliability, within a range of from one to two prob-

able errors. *c.* In five of the six tests there is marked over-estimation when the correlation of the halves of test *A* is substituted in the formula. The same is true of the predictions based upon the correlation of the odd and even elements. Inasmuch as these two sets of predictions are the ones of most practical import, the fact that they are far too high is interesting. It would seem that one should be very cautious in estimating the result to be expected with repeating a test from a value secured under the relatively constant conditions that prevail at a single testing. *d.* In three of these tests the correlation of the first half with the second furnishes the more accurate basis for prediction, while in the other three the odd-even correlation leads to better estimations of reliability. *e.* The results from the tests for pitch, consonance and rhythm show that all three might be decreased in length at least one-half with no loss in reliability."

The present writer has reported certain further data concerning the relation of the Spearman-Brown formula and the Seashore tests. His method has been to cut each test in half, find the odd-even reliabilities of these halves, step them up by the formula, and then compare the results with the previously determined reliabilities of the entire tests. Thus, odd-even reliabilities of the upper five rows, and similarly of the lower five rows, were figured for the pitch, intensity, and time records. These values were stepped up, and were then compared with the top-bottom reliabilities of the entire tests. As has been previously mentioned, the intensity test was split in a second way. It is so arranged that face A is roughly comparable to face B. It was, then, a simple matter to find the reliabilities of face A and face B, and to compare their stepped-up values with the correlation between the faces. A somewhat similar split of the time test has been mentioned

as possible. Columns A, C, E, G, and I make up a half section roughly comparable to Columns B, D, F, H, and J. The augmented reliabilities of these separate sections were compared with the intercorrelations of the sections. These data make it seem "that the Spearman-Brown formula functions as perfectly as can be hoped when reliabilities on groups agree no better than they do" (14).

Elsewhere the writer has given his reasons for believing that the split-half method of obtaining reliabilities is a better one than the principle of retesting. Lanier (43, 44), although mentioning the lack of proper testing conditions common to second renditions, appeared to feel that retest reliabilities are better. In correspondence with the writer, he has stated that it is a moot question.

Attempts to Raise Reliabilities

Consonance. In an attempt to raise its reliability, the writer decided to treat the Sense of Consonance Test as if it were a preference test. This is what a majority of subjects make it, regardless of its complicated directions. First of all, the subjects' responses were tabulated. Only those items were kept in the scoring key which showed significant differences, that is, were predominantly liked or disliked. It was decided that, for a percentage of likes or dislikes to be significant, the D/σ_D must equal 3.[12] For the population used, this meant that the percentage of difference must

[12]The writer gratefully acknowledges the statistical aid of Q. McNemar and E. A. McAnulty in this connection.

be at least 63-37. That is, at least 63% of the votes for a certain item were to be for "like" (or "dislike") before that item was kept in the test. This caused the elimination of several items. The subjects' ballots formed the criteria of correctness.

Sad to say, the odd-even reliability of the consonance test was even lower when scored in this new manner (.30±.05) than it was when the usual scoring sheet was employed (.36±.04). The new procedure failed to add benefit here.[18]

Kwalwasser Tests. When this procedure was applied to the Kwalwasser tests, the reliabilities were slightly enhanced. The value for melody changed from .26±.05 to .47±.05. That of harmony, which had been .17±.05, became .61±.04. Other subjects were given the tests to ascertain the stability of this method of procedure. Harmony scores, which showed a reliability of .21±.06 when the Kwalwasser scoring sheet was used now gave a value of .51±.05. Similarly, melody scores which had a reliability of .42±.05 now became .49±.05. It has thus proved possible to raise slightly the odd-even reliabilities of the Kwalwasser tests. They are still low, however, approximating those found for the Seashore time test.

Rhythm. It has been noted above that the rhythm test's reliability is extremely low. In a succeeding section it will be shown that the Stanford median values do not check with the original Iowa norms. The writer decided to change the phonograph speed from 78 r.

[18]Reported in part at the 1929 meeting of the Western Psychological Association.

p.m. (the speed used for all the tests with the exception of Sense of Time) to that of 60 r.p.m. (the speed employed for the time test) in an effort to help matters. The reliability, however, was even lower (a drop from .47±.04 to .37±.04 on a comparable group).

Intensity. All the testing reported by the present writer has been done on old-style phonographs comparable to the instruments employed by Seashore in the original work of standardization. With the exception of one study, to be reported later, this is true of all the studies of others. At least, one can find no statements to the contrary. Largely to find the effect on the mean and median scores, the writer varied this procedure with 108 university students by giving the intensity test on the Panatrope[14] at its maximum intensity. The odd-even reliability fell from the values previously reported to .42±.05 for this new group.

R. C. Larson's Data. Mrs. Larson (47) has assembled data in which retests were given immediately following the first rendition of the tests. Her subjects were two groups of university students—one with a population of 200, and the other of 100. Her values were as follows: pitch, .80±.02, .89±.01; intensity, .75±.02, .78±.03; time, .68±.03, .73±.03; consonance, .71±.02, .74±.02; tonal memory, .92±.01, .95 ±.01; rhythm, .68±.02, .72±.03. These values are considerably higher than the correlations reported so far. This may be due to a better type of technique employed or to a factor unique in this variety of reli-

[14]The writer is grateful to the Sherman-Clay Co., of Palo Alto, California, for the loan of the Panatrope.

ability coefficient. The present writer is inclined to the belief that there is grave danger of at least a small memory carryover which will be sufficient to make the correlations spuriously high. For this reason he does *not* accept these values as reliability coefficients in the strict sense of the term.

Salisbury-Smith Data. Another experiment in which retests were immediately given was carried on by Salisbury and Smith (65). The values for freshmen normal-school students were: pitch, .93±.01; intensity, .90±.01; time, .83±.02; consonance, .70±.03; memory, .94±.01; rhythm, .83±.02.

These experimenters also modified the pitch test so that the items were scaled from less to more difficult. The modified and original forms of the pitch test were found to intercorrelate with a value of .90±.04. This was presumably of the nature of an immediate retest, and, if so, suffers from the error mentioned above.

Lanier's Data. With a group of 130 subjects, the Seashore time test was found by Lanier (45) to have a retest reliability of .45. However, when a similar study was made of 20 male sophomore students in a class in experimental psychology in New York University, the retest value fell to .39, although the directions were carefully followed. With the 130 subjects, rhythm retested at .50. With the 20, it fell to .42. Lanier is now working on six different approaches to the measurement of time intervals.

Reliabilities of Immediate Retests. The opinion has been expressed that a retest given immediately following the first rendition of a test should have a higher reliability than the initial test. The writer's

data show little difference in the magnitudes of the reliabilities. What might be gained in "test wisdom" and the like on the retest is apparently compensated for by the ennui of the procedure. The following odd-even reliabilities were found for certain Stanford groups—initial pitch value, .72±.03; pitch retest, .69 ±.03; initial pitch value, .68±.03; pitch retest, .67± .03; initial intensity value, .62±.03; intensity retest, .64±.03.

In her data on immediate retesting R. C. Larson (47) found little difference in the means of the two testings. This fact would seem to work in well with the present writer's findings reported above.

INTERCORRELATIONS

In a discussion of a test battery it is important to learn the intercorrelations of the test scores. C. E. Seashore (67), who has devised the majority of the tests in question, has reported a correlation of .17± .04 between the pitch and time tests. In an article which he and Mount (72) have written the findings covered several years. The 1911 data reveal that pitch discrimination correlated with intensity to the extent of .13±.05, and with consonance .33±.05. The consonance-intensity correlation gave a value of −.03±.05 (Spearman *R* employed).

From their 1916 data (Pearson product-moment *r* used), the correlations shown in Table 1 were assembled. From this table one might conclude that the only serious overlapping among these tests is between pitch discrimination and tonal memory.

TABLE 1
(From Seashore and Mount)

	Consonance	Intensity	Time	Tonal memory
Pitch	.21±.04	.09±.04	.17±.04	.52±.03
Consonance		.11±.04	.07±.05	.23±.04
Intensity			.15±.04	.26±.04
Time				.17±.03

TABLE 2
(From Weaver)

	Consonance	Intensity	Time	Tonal memory
Pitch	.78±.03	.32±.06	.30±.06	.52±.05
Consonance		.24±.07	.48±.05	.75±.03
Intensity			.23±.07	.20±.07
Time				.28±.06

TABLE 3
(From R. H. Seashore)

	Time	Tonal memory	Rhythm
Pitch	.37±.08	.64±.06	.32±.09
Time		.57±.06	.52±.07
Tonal memory			.56±.06

With an early form of the consonance test, Malmberg (50) found that it correlated .18 with pitch, and .34 with tonal memory.

A study of Weaver (85) is reported in Table 2. Weaver's correlations are somewhat larger than those of Seashore and Mount. Consonance is especially at variance. This test is so unreliable, however, and its directions are so misleading, practically any correlation might be expected.

A correlation between scores on the rhythm and time tests given at the State University of Iowa was reported as —.02±.04 by R. H. Seashore (73). On

TABLE 4
(From R. H. Seashore)

1. Rhythm	$r_{12.3}$	.30
2. Time	$r_{13.2}$	.39
3. Tonal memory	$r_{23.1}$	.38
1. Rhythm	$r_{12.3}$	.50
2. Tonal memory	$r_{13.2}$	.04
3. Pitch	$r_{23.1}$	.58

TABLE 5
(From Brown)

	Consonance	Intensity	Time	Tonal memory	Rhythm
Pitch	.29±.06	.25±.06	.20±.06	.52±.05	.16±.06
Consonance		−.04±.07	.09±.07	.22±.06	.09±.07
Intensity			.40±.06	.00±.07	.23±.06
Time				.20±.06	.00±.07
Tonal memory					.33±.06

other subjects this same writer found the correlations given in Table 3. By the use of the partial correlation method on portions of the data, further analyses were obtained (Table 4). These correlations are *not in keeping with those of previous studies nor with* the data to be reported later. The differences are possibly due in part to the technique employed and to the small number of subjects. All subjects who scored below 50 percentile were retested and given the grades made on the second trial. During this same study Seashore selected from a class the subjects who made the 13 highest and the 13 lowest scores on the rhythm test, and correlated their scores with those on tonal memory. This technique gave the high correlation of .91±.02.

Brown (7), in a study of high-school juniors and seniors, found the correlations reported in Table 5.

McGinnis (51) has reported a study made by McCarthy on school children, junior-high, and college students. Intensity was found to intercorrelate with pitch from .09 to .38. Consonance-pitch intercorrelations ranged from .21 to .48, and consonance-intensity from .10 to .16. The same intercorrelations on McGinnis' preschool children were .49, .62, and .26, respectively (rank coefficients).

Rhythm was found by De Graff (11) to correlate with tonal memory to the extent of .46, with intensity .07, and with time .25. By the use of partial correlations, these became: rhythm and time with intensity constant, .25; rhythm and intensity with time constant, .00.

Hollingworth (32) has studied a group of 49 fifth-grade children (median age 10 years 4 months) who have IQ's above 135. Table 6 displays the results of Seashore intercorrelations. The relatively high correlation between pitch and time is the striking variant here.

In his book, *Tests and Measurements in Music,* Kwalwasser (40) has commented as follows: "The correlations between scores earned on time and rhythm and intensity and rhythm are very low. It was found, however, that the correlation between tonal memory

TABLE 6
(From Hollingworth)

	Consonance	*Intensity*	*Time*	*Tonal memory*
Pitch	$.34 \pm .09$	$.15 \pm .09$	$.51 \pm .07$	$.46 \pm .08$
Consonance		$.09 \pm .10$	$.47 \pm .08$	$.34 \pm .09$
Intensity			$.08 \pm .10$	$-.09 \pm .10$
Time				$.25 \pm .09$

TABLE 7
(From Salisbury and Smith 1926)

	Consonance	Intensity	Time	Tonal memory	Rhythm
Pitch	.34±.05	.32±.05	.38±.05	.49±.05	.20±.06
Consonance		.31±.06	.32±.05	.42±.05	.06±.06
Intensity			.43±.05	.41±.05	.25±.06
Time				.39±.05	.42±.05
Tonal memory					.51±.04

TABLE 8
(From Salisbury and Smith 1927)

	Consonance	Time	Tonal memory	Rhythm
Pitch	.30±.05	.45±.05	.60±.04	.43±.05
Consonance		.32±.05	.26±.05	.09±.06
Time			.39±.05	.47±.05
Tonal memory				.51±.04

and rhythm is rather high (+.49). This is the highest correlation to be found between any two tests of the Seashore battery." The present writer does not know upon what data Kwalwasser has based these statements.

Scores by freshmen normal-school students were found by Salisbury and Smith (65) to have intercorrelations as shown in Table 7.

These same experimenters gave four of the battery again the following year to 144 entering freshmen who were taking music. The pitch test was modified so that the items progressed from less difficult to more difficult.

Data on Stanford students are presented in Table 9.[16]

Gaw (23) has worked over the original Iowa

[16]Reported in part at the 1927 meeting of the Western Psychological Association.

TABLE 9
(From Farnsworth)

	Consonance	Intensity	Time	Tonal memory	Rhythm	Melody	Harmony
Pitch	.36±.04	.15±.05	.11±.05	.50±.04	.10±.05	.29±.05	.24±.05
Consonance		.10±.05	—.09±.05	.39±.04	.04±.05	.25±.05	.28±.05
Intensity			.11±.05	.12±.05	.17±.05	.02±.05	.02±.05
Time				.18±.05	.23±.05	.15±.05	.02±.05
Tonal memory					.38±.04	.30±.05	.37±.04
Rhythm						.09±.05	.15±.05
Melody							.24±.05

TABLE 10
(From Gaw)

	Consonance	Intensity	Time	Tonal memory
Pitch	.24±.03	.82±.01	.54±.02	.77±.01
Consonance		.08±.03	.02±.03	.17±.03
Intensity			.95±.00	.63±.02
Time				.73±.02

TABLE 11
(From Broom)

	Consonance	Intensity	Time	Tonal memory	Rhythm
Pitch	.89±.01	.92±.01	.90±.01	.88±.01	.90±.01
Consonance		.87±.02	.86±.02	.94±.01	.89±.01
Intensity			.96±.01	.89±.01	.86±.02
Time				.88±.01	.83±.02
Tonal memory					.90±.01

standardization data for approximately 490 fifth-graders. The intercorrelations do not greatly resemble those of adults.

The subjects of an intercorrelation experiment by Broom (6) were ninth-graders.

Peacock (59) found low intercorrelations for Grades 6, 7, and 8. The Seashore tests made up the battery studied.

R. C. Larson (47) has reported intercorrelations on the Seashore tests. In Table 12 the values for the eighth-graders will appear above those for the adults.

The same researcher has run intercorrelations for the seventh and fifth grades. In Table 12*a* the values for the fifth-graders will appear above those for the seventh-graders.

Unpublished data worked out by Norval Church under the guidance of the present writer gave results for fifth- and eighth-graders which are quite compar-

TABLE 12
(From Larson)

	Consonance	Intensity	Time	Tonal memory	Rhythm
Pitch	.47±.05 .32±.06	.45±.05 .33±.06	.38±.06 .45±.05	.71±.03 .49±.05	.54±.05 .30±.06
Consonance		.11±.04 .08±.07	.16±.07 .25±.06	.30±.06 .37±.06	.30±.06 .15±.07
Intensity			.33±.06 .30±.06	.37±.06 .24±.06	.23±.06 .21±.06
Time				.47±.05 .24±.06	.40±.06 .55±.05
Tonal memory					.52±.05 .37±.06

TABLE 12*a*
(From Larson)

	Consonance	Intensity	Time	Tonal memory	Rhythm
Pitch	.46±.05 .26±.08	.15±.07 .49±.07	.25±.06 .33±.08	.47±.05 .56±.06	.23±.06 .16±.09
Consonance		.16±.07 .18±.09	.20±.07 .15±.09	.19±.07 .29±.08	.16±.07 .23±.09
Intensity			.38±.06 .39±.08	.28±.06 .69±.05	.46±.05 .29±.08
Time				.46±.05 —.02±.09	.36±.06 .26±.08
Tonal memory					.41±.06 .11±.09

TABLE 13
(From Church, Farnsworth)

	Consonance	Intensity	Time	Tonal memory	Rhythm	Melody	Harmony
Pitch	.18±.05 .18±.05	.21±.05 .25±.05	.28±.05 .20±.05	.30±.05 .33±.04	.32±.05 .07±.05	.26±.05 .18±.05	.04±.05 .14±.05
Consonance		.19±.05 .22±.05	.30±.05 .15±.05	.21±.05 .33±.04	.07±.05 .20±.05	.29±.05 .13±.05	.17±.05 .24±.05
Intensity			.29±.05 .45±.04	.10±.05 .42±.04	.19±.05 .13±.05	.31±.05 .14±.05	.08±.05 .20±.05
Time				.23±.05 .29±.04	.38±.04 .22±.05	.11±.05 .11±.05	.25±.05 .20±.05
Tonal memory					.13±.05 .61±.03	.44±.04 .15±.05	.10±.05 .32±.04
Rhythm						.19±.05 .11±.05	.17±.05 .21±.05
Melody							.40±.04 .29±.04

able to those for adults. In Table 13 the fifth-grade results are above the eighth.

At all ages there appear to be positive correlations among the members of the Seashore-Kwalwasser battery. The Gaw and Broom data indicate that the correlations are occasionally large; the remainder of the studies would seem to show that the correlations are small. The writer has no rationalization to offer in explaining away this discrepancy.[16]

STANDARDIZATION[17]

Well-standardized tests, if given under the proper conditions to large groups of unselected subjects,

[16]In a recent article by McCarthy (50*a*), intercorrelations were reported which were very similar to those found by the present writer.

[17]Just before this monograph went to press, R. C. Larson (47) issued new norms.

should yield median scores which equal the fiftieth percentile ratings of the criterion groups. The eight "music" tests in question have been checked in this manner by a number of different workers.

R. H. Seashore (73) has reported mean percentiles on 44 university subjects. The method of averaging percentiles does not give quite the result obtained by averaging raw scores and then ascertaining the median, but the end effect is at least somewhat similar. Seashore's results were as follows: pitch, 47.2; time, 58.7; 51.3 for tonal memory; and 72.2 for rhythm. The large variation from 50 percentile in the case of rhythm should be noted. In the data to be reported from this point on, the median or mean scores will be translated into their *approximate* percentiles on the Seashore norms.

Lanier (43) has recorded mean scores on two presentations of the Seashore battery. His percentile values were: pitch, 11, 12; intensity, 14, 27; time, 42, 33; consonance, 36, 34; tonal memory, 21, 14; rhythm, 35, 24 (N=106 to 109). These values obviously vary considerably from 50. This fact may be accounted for by at least four hypotheses. (*a*) The Seashore norms may be incorrect. (Data to be reported subsequently indicate that this theory will *not* explain the peculiarity of Lanier's data.) (*b*) The directions or the technique of procedure may not have been followed accurately. (The present writer does not feel that this is a tenable hypothesis.) (*c*) The subjects may have been below the mean of Seashore's criterion groups in the test abilities in question. (The Lanier

subjects were students of all college classes of the Middle Tennessee State Teachers College of Murfreesboro, Tennessee.) (*d*) The subjects may not have paid as careful attention as did the members of the Seashore criterion groups. (Lanier has mentioned a lack of attention on occasion.) An occasional lack of attention is an observation common to most experiments. However, it is practically impossible to equate the amount of attention exhibited in the various studies.

Brown (7) has also reported mean scores on two presentations of the Seashore battery. (It should be recalled that the median is the fiftieth percentile. In large distributions which follow the normal curve, the mean should approximate the median.) Brown's percentile values were: pitch, 32, 10; intensity, 4, 31; time, 44, 24; consonance, 90, 36; tonal memory, 31, 38; and rhythm, 47, 47 (if figured on Seashore's adult norms). The subjects were junior- and senior-high-school students, so might fall between eighth-graders and adults (college sophomores). The values on the eighth-grade norms were: pitch, 51, 23; intensity, 11, 64; time, 81, 60; consonance, 92, 49; tonal memory, 50, 60; rhythm, 70, 70 (N=89 to 93). Note the extreme variability found in retesting. This is partly due to the fact that an extremely slight alteration of the raw score will frequently cause a sizable change in the percentile standing. Note also that the rhythm percentile is considerably over 50.

Weaver (85) has studied the test scores of 94 subjects. His mean scores approximated these percentiles: pitch, 31; intensity, 52; time, 55; consonance, 52; and tonal memory, 41.

As reported by Peterson and Lanier (62), white students from Peabody made scores which were equivalent to 14 percentile on pitch on the first presentation (N=330), and 27 on the retest (N=89). The corresponding values for consonance were 29 and 47 (N=330 and 89). The increase in score on the retest may be due in part to selective factors. A similar study of Fisk colored students gave values of 19 (N=274) and 28 (N=286) for pitch, and 47 (N=274) and 42 (N=286) for consonance. Further work was done on white subjects from Middle Tennessee State Teachers College (N=372 to 386) and on colored from the Agricultural and Industrial Normal College of Nashville (N=288 to 296). The median percentiles for Whites and Negroes respectively were: pitch, 8, 4; intensity, 28, 7; time 40, 16; consonance, 40, 23; tonal memory, 14, 3; rhythm, 24, 24. As was the case with Lanier's work reported above, the percentiles are surprisingly low. The hypotheses invoked above may be reinvoked here. The significance of the racial differences will be discussed later.

D. L. Larson (46) has studied the consonance scores of 35 university students who were rated on questionnaires as "musical." They gave a mean score which approximated 85 percentile. Thirty-five untrained subjects gave a value of 23. Further trained observers (N=35) who were tested individually scored 95 percentile, while a "corresponding untrained group" averaged 14. In another study of groups of 35 subjects, the following data were assembled. Trained subjects given the consonance test in a group averaged

86 on both the first rendition and the retest. Individual testing and retesting gave 96 and 93. Untrained subjects tested in a group gave values of 27 on the first rendition and 20 on the retest. Individual testing and retesting gave 18 and 11. The mean for 132 cases approximated 43 percentile; for 150 cases, 40 percentile.

Thirty university subjects, on whom Heinlein (29) experimented, gave a median percentile of 43 on the pitch test. In the course of another study, Heinlein (27) tested and retested groups of 30 and 35 on the consonance test. The mean scores gave the following approximate percentiles (figured by the present writer and not entirely accurate): 33, 36, 53, 33. A group of 107 male subjects was found to have a mean score which, when translated into percentile terms, gave a value of 39.

Poorman has conducted an unpublished study of the rhythm test presented at its normal speed, 78 r.p.m., and at 60 r.p.m. Eighty-two university students were given the test at speed 60, then at speed 78, and two weeks later at speed 78. The approximate percentiles of the mean scores were 48, 66, and 74, respectively. Seventy-six students were given the test at speed 78, then at 60, and two weeks later at 78 again. The approximate percentiles of the mean scores were 56, 38, and 76.

Barnes (1) has given the six Seashore tests to 200 students of Maryville College. He has stated that the average score equalled 39 percentile.

Two groups of normal-school students were tested

by Salisbury and Smith (65). Their mean scores approximated the following percentiles: pitch, 35, 45; intensity, 42; time, 39, 43; consonance, 65, 51; memory, 40, 50; rhythm, 62, 70.

One hundred eighty-seven summer-school students of the San Francisco Teachers College were found by Gaw (23) to score considerably below the Seashore standards. The percentiles for pitch, intensity, and memory were 32, 29, and 35, respectively. Concerning her results, Gaw has written:

> "The reason for the differences in all three tests, assuming that the San Francisco teachers should have an average score equal to the fiftieth percentile of Seashore, is *obscured because the methods of evaluation are not* exactly identical. It is possible that a true difference between the Seashore group and the San Francisco group exists. The difference is made clearer by a consideration of the measures of spread, since Seashore's spread in terms of standard deviation can be estimated roughly from his percentile tables .... by finding the *scores nearest to the sixteenth percentile and the eighty-*fourth percentile and taking half the difference. Estimating Seashore's standard deviations in that manner, they are:

	Score 16th percentile	Score 84th percentile	One-half difference or estimated sigma
Pitch	71	86	7.5
Intensity	81	93.5	6.25
Memory	57	88	16.25

> Although the San Francisco means for Pitch and Intensity are lower than Seashore's means, they show a considerably greater deviation. In the case of the Memory test the whole distribution of scores is definitely lower in the San Francisco results."

The present writer has determined medians and means for 200 Stanford subjects. The approximate

percentiles are here presented, in the order—median, mean: pitch, 63, 54; intensity, 45, 43; consonance, 46, 51; time, 54, 50; memory, 55, 57; rhythm, 84, 77; melody, 68, 69; harmony, 69, 72.[18]

R. C. Larson (47) has very recently developed new norms for the Seashore battery. Her median adult scores were found to approximate the following percentiles on the old norms: pitch, 50; intensity, 58; time, 48; consonance, 51; tonal memory, 50; and rhythm, 35. A group of eighth-graders gave these values: pitch, 55; intensity, 56; time, 50; consonance, 54; memory, 45; rhythm, 45. Data have also been assembled for fifth-graders as follows: pitch, 51; intensity, 55; time, 54; consonance, 54; memory, 50; rhythm, 45.

Eighty-two juniors and seniors from San Diego State Teachers College (Broom, 6) gave the following average percentiles: pitch, 49; intensity, 46; time, 53; consonance, 50; memory, 51; and rhythm, 67. One hundred two ninth-grade pupils gave mean percentiles *(on the adult and eighth-grade Seashore norms)* as follows: pitch, 9, 21; intensity, 2, 5; time, 21, 57; consonance, 41, 65; memory, 42, 66; and rhythm, 55, 75. It can be seen that the ninth-graders were apparently not typical. Both groups were too high in rhythm.

In a study of 191 white and 191 colored fifth-graders in the Gary, Indiana, schools, Lenoir (49) found mean scores which the writer has translated into percentiles. In each instance the values for the colored

[18]Reported in part at the 1927 meeting of the Western Psychological Association.

students are given first. The percentiles are: pitch, 43, 48; intensity, 38, 47; consonance, 56, 66; memory, 56, 55; time, 62, 50; rhythm, 55, 33.

Davenport (9, 10) has used white, black (colored), and brown (mixed black and white) farmers as subjects in a study of race-crossing in Jamaica. On the pitch test the percentile equivalents of the mean scores were: white, 17; black, 26; brown, 32. On the rhythm tests the values were: whites, 69, and blacks, 92. The browns ranged from 0 to 100 (mean score not given).

Klauer (37) has studied the effects of training on the rhythm test. The percentile rank of the mean score of her training group was considerably above that of her control group.

Kwalwasser's study (40) stands practically alone in finding the rhythm standardization correct—at least at the adult level. The median percentiles for two groups were: fifth grade, 65, 55; eighth grade, 49, 47; adult, 48, 46.

A study by Fracker and Howard (16) is difficult to interpret in percentile terms since the scores were arranged in intervals of 5 and modes were figured instead of medians or means. To quote:

> "It will be seen that our distributions are fairly regular and that the characteristic values correspond quite closely with those that have been found with the much larger material measured by Professor Seashore. In one test only, that of consonance, the scores made at Arkansas exceed the Seashore mode."

Windhorst (87) has developed sixth-grade norms for all the Seashore tests. She found that the rhythm norms were not comparable to values estimated from

the standard fifth- and eighth-grade norms. Her values were too high, as were those in the present writer's study.

Subjects from sixth, seventh, and eighth grades were tested by Gray and Bingham (25). Two hundred fifty-eight were colored and 219 were white. Since their medians should fall in between the Seashore fifth- and eighth-grade norms, percentiles will be given on both in the order—colored eighth grade, colored fifth grade, white eighth grade, white fifth grade. The percentiles were: pitch, 9, 23, 21, 46; intensity, 7, 36, 37, 73; time, 27, 58, 46, 76; consonance, 44, 66, 38, 61; memory, 7, 33, 36, 66.

Under the guidance of the present writer, fifth- and eighth-grade students of San José, California, were recently tested by Norval Church. The median percentiles are here reported in the order—fifth grade, eighth grade: pitch, 49, 51; intensity, 48, 46; time, 41, 50; consonance, 50, 49; memory, 41, 59; rhythm, 51, 24; melody, 60, 75; harmony, 41, 61. It should be noted that the melody and harmony tests seem poorly standardized for these grades. The writer has no explanation for the low rhythm score of the eighth grade.

In summarizing, one can state that the Seashore tests (with the possible exception of rhythm) frequently give the median percentiles they should (as shown by the studies of the writer and others). Some researchers, however, have presented different results. The Kwalwasser tests do not appear to be well standardized.

AGE DIFFERENCES

The Seashore tests have three distinct norms—for fifth-grade pupils, eighth-grade pupils, and adults. Seashore (67) has justified this by the introduction of the notion of physiological and cognitive limits. All tests, he has thought, aim to disclose the absolute, or physiological, limits in an ability. They fail to the extent that they become monotonous and uninteresting, or perhaps unintelligible. What they test, then, is the cognitive limit. Thus, there is more chance of tapping the adults' physiological limits than those of eighth-grade pupils, and more likelihood of obtaining the physiological limits of eighth- than of fifth-grade pupils. This is a tenable hypothesis. It may be found, however, that physiological limits change appreciably with age.

The statement by Smith (74) that "the amount in favor of the adult shown in all group statistics is amply accounted for by the difficulty in making a reliable test on the young and by their lack of information" seems hard to reconcile with the fact that in intelligence tests the average mental age of the eighth-grade pupil is found to be almost, if not quite, that of an adult.

The work of Hollingworth (32), to be discussed again later, also tells against the second portion of Smith's assumption. Forty-nine children whose IQ's ranged from 135 to 190 with a median of 153 (Stanford-Binet) gave about the same values on five of the Seashore tests as do normal children. And this occurred in spite of the fact that "their median

TABLE 14

TYPICAL PERCENTILES IN THE SEASHORE TESTS FOR THE THREE AGE GROUPS

% re-sponses correct	Pitch			Intensity			Time			Rhythm			Consonance			Memory		
	5	8	Ad.	5	8	Ad.	5	8	Ad.	5	8	Ad.	5	8	Ad.	5	8	Ad.
55	21	7	3	8			17			11	4		27	11	4	59	27	14
70	59	31	15	38	8	3	75	45	13	76	50	23	83	66	56	82	64	41
80	87	63	45	73	36	14	97	90	61	97	88	77	98	97	97	94	84	65
95	100	100	100	100	98	92	100	100	100	100	100	100	100	100	100	100	99	96

intellectual ability was, therefore, at least that of the average adult."

In view of the norms established by Seashore for his rhythm test, this statement by De Graf (11) is surprising:

> "It is usually believed that age influences rhythmic discrimination; that the older the person is, the better his rhythmic discrimination will be, but such is not the case, as a glance at the distribution table (his own) will show the distribution curve of the three age groups, namely, fifth-grade, eighth-grade, and adult, reveals an unusual degree of similarity. Although the adult has the advantage of maturity, enabling him to comprehend and concentrate on the test, his score on the rhythm test is only slightly higher than that of the fifth- and eighth-grade school child."

In discussing his two tests, Kwalwasser has stated: "It will be observed that the norms make no allowance for differences in either age or training, which means that these factors do not make for significant differences in scores" (40). In the present writer's study the fifth-graders displayed median percentiles on the melody and harmony tests of 60, 41; the eighth-graders, 75, 61; and the adults, 68, 69. No definite conclusion can be made from these data, however, since the tests are quite unreliable, and a change of one point in raw score occasionally alters the percentile ranking by as much as ten points.

At the present writing one can state only that, while age differences are to be found in the Seashore battery, the causative factors underlying these phenomena are not obvious.

Relations to General Intelligence

In his *Psychology of Musical Talent,* C. E. Seashore has reported that not one of his battery of tests correlates to any great extent with estimated intelligence. Smith (74), a Seashore student, had previously found appreciable correlations between pitch discrimination and estimated intelligence; but the Seashore pitch record had not been used.

Weaver (85) ran correlations between a portion of the Seashore battery and the Army Alpha test, with the following results: pitch, .35±.06; intensity, .24±.06; time, .12±.07; consonance, .06±.07; and tonal memory, .26±.06.

In studying male students at the State University of Iowa, R. H. Seashore (73) has found a correlation of .10±.09 between percentile rank in the Iowa Qualifying Examination and rank in the rhythm test. This same intelligence test was employed by Travis and Davis (83) and found to correlate with pitch in the case of three groups of subjects: —.01±.07, .02±.07, and .05±.06. They also stated that the qualifying examination correlated .07±.06, .11±.07, and .02±.06 with intensity; and —.08±.06, .07±.07, and .04±.05 with memory.

Peterson (61) has reported correlations with the Otis test ranging from .04 for consonance (on white students) to .45 for tonal memory (on Negro students).

In his book, *Tests and Measurements in Music,* Kwalwasser (40) has given the following correlations with mental age (the "intelligence" test is not specified): pitch, .20; intensity, .25; time, .21; consonance,

.01; memory, .02; rhythm, .01. The *P.E.*'s are all of .04.

Beach (3) has found a correlation between the Seashore tests and Army Alpha on 450 college freshmen of "less than .10." Barnes (1), in a study of 200 college students, has stated that there was "very little" correlation between the tests and intelligence.

Fracker and Howard (16) have studied the relationships between intelligence tests—the Otis Self-Administering (higher examination) and Army Alpha —and the Seashore tests. Two hundred thirty freshman, sophomore, and junior students of the University of Arkansas served as subjects. The coefficients were: pitch, .32±.04; intensity, .01±.05; time, .13±.04; consonance, .09±.04; memory, .10±.04; and rhythm, .12 ±.04.

The work of Hollingworth (32) has already been discussed in part. Her 49 "intellectually gifted" child subjects gave mean percentiles on the fifth-grade norms of: pitch, 47; intensity, 50; time, 58; consonance, 48; and memory, 52. Hollingworth concluded:

> "Above the level of intelligence required to understand and execute the directions for taking the Seashore tests (mental age of about ten years), performance in pitch discrimination, perception of intensity, perception of consonance, and tonal memory is not symptomatic of intellectual endowment."

Highsmith (31) has ascertained the relationship between certain of the Seashore battery and a combined intelligence rating made from the Terman Group Test and the Thurstone Psychological Examination. The values were found to be: pitch .58±.06; intensity, .35

±.08; time, .39±.08; consonance, —.14±.09; memory, .30±.08; average Seashore, .36±.08. The subjects were students of a school of music. He has concluded:

> "There is apparently very little measured by the Seashore tests which is not measured also by the intelligence test."

Salisbury and Smith (65) have carried on a similar piece of work with normal-school students. The coefficients of correlation were: pitch, .31±.05; intensity, .15±.06; time, .30±.06; consonance, .01±.06; memory, .24±.06; rhythm, .02±.06. A selected group who were taking music courses gave values of: pitch, .39 ±.05; time, .49±.04; consonance, .38±.05; memory, .34±.05; rhythm, .24±.05.

Baugh (2) has studied the "development of rhythm perception through training." The IQ's of her fifteen subjects showed no relationship to the rhythm scores.

Gray and Bingham (25) have added together the Seashore test scores of their subjects, and by this method created what they called a *total music score.* The results of the Otis Group Advanced Examination Form A (and perhaps the Stanford Achievement test) made up an *index of brightness.* These intercorrelated to the extent of .58±.09 for colored boys, .53±.11 for colored girls, .70±.05 for white boys, and .68±.05 for white girls. The subjects were from sixth, seventh, and eighth grades.

Brown (7) has averaged the Seashore scores of his junior- and senior-high-school subjects, and has found a correlation of .24 with intelligence ratings.

The present writer has correlated scores on the Seashore and Kwalwasser tests with Thorndike scores.

The results were: pitch, .14±.05; intensity, .11±.05; time, .10±.06; consonance, —.38±.04 (an elimination of four extreme cases changes this correlation to .02 ±.05); memory, .11±.05; rhythm, .17±.05; melody, —.08±.06; and harmony, .22±.05.[19] [20]

There are at least three possible approaches to the problem of the relation of these test scores to "intelligence." One is to run correlations between the two series, as has been done by several writers. A second is to see whether or not subjects who make high scores on "intelligence" tests also rank well on the "music" tests. Such was the work of Hollingworth cited above. The third obvious method is to ascertain whether the high scorers on the "music" tests are better than average on the "intelligence" tests. A study of the writer of this sort is listed below. Mean Thorndike scores were found for two groups: those with "music" percentiles from 25 to 75, and those whose values were from 75 to 100 percentile. From the resulting data (Table 15) it can be readily seen that there are no consistent relationships between the test battery and Thorndike scores.

From a survey of the studies here reported, the writer concludes that the "intelligence" tests and the Seashore-Kwalwasser battery test very different variables in human behavior. As Hollingworth suggests, there is probably a minimal mental age which must be

---

[19]A recent study by R. C. Larson (47) (received as this monograph was ready for press) offers data on this topic. Correlations between "intelligence" and Seashore scores were low.

[20]The Stanford data were reported in part at the 1927 Western Psychological Association.

TABLE 15

	25 to 75	Sigma	No.	75 to 100	Sigma	No.	Chances in 100	D/σ_D
				Pitch				
M.Th.*	(80.50)	11.95	162	(80.24)	11.60	117	57	.18
				Intensity				
M.Th.	(77.40)	12.11	185	(80.70)	12.70	89	98	2.05
				Time				
M.Th.	(79.90)	12.34	104	(80.11)	12.13	67	54	.10
				Tonal memory				
M.Th.	(77.53)	12.65	125	(80.50)	13.03	62	93	1.50
				Rhythm				
M.Th.	(80.94)	9.95	124	(77.27)	11.04	84	99	2.40
				Consonance				
M.Th.	(79.26)	12.24	107	(78.12)	12.84	62	72	.57
				Melody				
M.Th.	(78.05)	12.22	162	(77.97)	12.28	144	52	.06
				Harmony				
M.Th.	(77.12)	13.04	183	(79.76)	11.41	119	96	1.86

*Mean Thorndike score.

possessed before one can score well on this battery. An appreciable correlation could without doubt be created if subjects were used who ranged from idiots to "gifted."[21 22] In ordinary populations, however, the correlations are low.[23]

Tetrad Differences. Spearman (75) has said: "the abilities to appreciate, for instance, the relations of pitch, loudness, and rhythm have extremely low inter-

[21]R. C. Larson (47) has reported a study in which the Seashore tests were given to Grades 1 to 5. There was no significant difference in scores between those of average IQ and those whose IQ's were above 110. Significant differences were found between the scores of those above average and those below average IQ.

[22]Information concerning the correlations between Seashore scores and "intelligence" computed by Peacock (59) was not available to the writer.

[23]In a study of Negroes and whites of the third, fourth, fifth, and sixth grades Streep (81*b*) found correlation values of .05 to .24 (N=136 or more) between rhythm scores and IQ. In a similar study of the consonance test the values ran from 0 to .25.

correlations; no more, in fact, than must be attributed to *g* alone." However, had Spearman chosen the memory record as one of his music tests, the intercorrelations would have been higher. The present writer has taken the Stanford table of intercorrelations for the Seashore tests only, and has figured the forty-five tetrads which the six tests make possible. Their median is .0286 with a $P.E._t$ of .0212. The reader may interpret these results as he sees fit.

Sex Differences

Prior to the construction of the pitch record, Smith (74) had discovered differences in pitch discrimination in favor of the elementary-school girl. He felt that this difference was entirely due to the aloofness of the boy whose social training was such as to make him regard music as girlish, and so to be avoided. About this time, Vance (84) had found that his 16 male subjects surpassed his 20 female subjects in the discrimination of pitch at every level in the register. The differences were the greatest at 64 and 2048 d.v.

C. E. Seashore (67) has stated in his *Psychology of Musical Talent* that there are no appreciable sex differences in pitch discrimination.

As reported by Kwalwasser (40), fifth-grade boys scored higher on the rhythm test than did the girls—67.74 to 65.83. Sixth-graders scored alike—69, while the boys of the eighth grade made higher scores—69.78 to 69.47. The adult males also made higher scores—74.12 to 73.81. Concerning these findings, Kwalwasser has said: "These median scores are approxi-

mately the same and reveal the sexes to be about equally endowed rhythmically." However, the same slight tendency for the males to give higher scores appears in the studies next to be discussed.

Table 16 shows the sex differences found by Peterson and Lanier (62) among white students of Middle Tennessee State Teachers College at Murfreesboro, and colored students of the Agricultural and Industrial Normal College of Nashville. Minus marks before the $D/P.E._D$ values indicate that the female subjects made higher scores. The unmarked indicate that the male subjects scored higher. The data are in terms of median differences.

Further work was done on 50 white male students and 248 white female students from Peabody, and 141 Negro men and 109 Negro women students from Fisk. On the consonance test the women scored higher with $D/P.E._D$ values of 3.8 (99 chances) for whites and 4.9 (100) for Negroes. On the pitch test the white men and Negro women were favored with values of 2.9 (97) and 3.4 (99), respectively.

Table 17 offers an interesting comparison. The writer here presents data on 200 Stanford students, 100 of each sex. The data are in terms of mean differences.[24]

It should be noted that the white subjects studied by Peterson and Lanier showed small sex differences whose directional variations were similar to those displayed by the Stanford subjects. The magnitudes were different, and were frequently not statistically reliable.

[24]Reported in part at the 9th International Congress of Psychology.

TABLE 16

Test	White No. M.	White score	No. F.	Score	$D/P.E._D$	C in 100	Negro No. M.	Negro score	No. F.	Score	$D/P.E._D$	C in 100
Pitch	125	61.70	249	63.70	—1.41	—85	94	58.23	194	57.62	.57	60
Intensity	123	85.59	250	84.08	2.14	92	95	82.08	201	72.84	5.78	100
Time	125	79.62	247	75.12	6.60	100	94	73.06	201	69.90	2.24	93
Consonance	121	65.48	265	66.86	—1.34	—82	87	62.87	201	64.15	—1.17	—78
Memory	119	53.75	263	56.70	—2.07	—92	88	45.62	201	42.08	—1.24	—80
Rhythm	121	70.21	264	70.00	.27	57	88	74.64	202	71.33	3.28	99

TABLE 17

Test	No. M.	Score	No. F.	Score	D/σ_D	Chances in 100
Pitch	100	80.4	100	82.9	—2.60	—99
Intensity	100	87.9	100	87.5	.53	69
Time	100	78.9	100	77.7	1.32	90
Consonance	100	67.2	100	70.8	—3.60	—100
Memory	100	73.0	100	80.4	—3.72	—100
Rhythm	100	81.2	100	78.6	2.50	99
Melody	100	25.9	100	26.2	— .65	—74
Harmony	100	24.9	100	25.9	—2.70	—100

The Negro subjects showed similar differences in all but the pitch tests.

Ten quite competent musicians who had previously listened to the Seashore battery were asked by the writer to guess the sex differences which he had already found. All but one declared that the girls should, on the average, score higher in pitch, memory, and consonance. The explanatory theories which were evolved were similar to the social-training idea of Smith (see above). The consensus of opinion seemed to be that boys are trained to be more self-reliant, and so would necessarily have more experience in judging time and rhythm. Having more mechanical opportunities, they would be trained to pay close attention to small differences in intensity. One musician decided that the girls should score higher in all the tests "since girls are naturally more musical."

Under the guidance of the present writer, Norval Church has tested fifth- and eighth-grade students of San José, California. The sex differences which were found are reported in Tables 18 and 19.

Although the sex differences of the eighth grade do not resemble those previously reported for college sub-

TABLE 18
FIFTH GRADE

Test	No. M.	Score	No. F.	Score	D/σ_D	Chances in 100
Pitch	99	63.5	103	65.1	.79	—79
Intensity	97	73.7	102	70.9	1.75	96
Time	97	61.2	101	60.0	.92	82
Consonance	99	59.4	103	62.4	2.31	—99
Memory	94	43.2	96	46.1	1.32	—90
Rhythm	97	64.6	97	65.4	.57	—72
Melody	99	24.1	97	25.0	1.63	—94
Harmony	98	21.2	96	22.2	1.67	—95

TABLE 19
EIGHTH GRADE

Test	No. M.	Score	No. F.	Score	D/σ_D	Chances in 100
Pitch	101	73.7	103	73.0	.12	54
Intensity	101	81.8	104	77.3	3.21	100
Time	101	71.9	102	68.8	2.59	99
Consonance	101	68.0	103	65.5	2.27	99
Memory	101	75.5	100	59.7	8.32	100
Rhythm	109	64.9	101	64.5	.34	64
Melody	100	26.9	102	26.3	1.20	88
Harmony	97	24.4	102	23.8	1.20	88

jects, the fifth-graders do somewhat follow the adult trends. However, it should be emphasized that few of the differences are statistically reliable. From his own observations, the writer hazards the theory that the sex differences found in a schoolroom will vary with the personal qualities of and the methods employed by the music teachers. In other words, the writer follows somewhat the ideas of the musicians mentioned above. Under such training conditions as have existed in the past, the girls have been slightly more interested in musical tones, and the boys in mechanical noises. Certain teachers may be able to

TABLE 20
PITCH ERRORS OF GROUP A

Columns Stimulus (d.v.)	A 30	B 23	C 17	D 12	E 8	F ½	G 1	H 2	I 3	J 5
% errors	1	6	4	5	6	49	53	48	36	18
	4	4	5	6	12	55	51	41	13	11
	4	4	6	7	13	60	24	35	24	9
	4	4	6	11	13	42	44	28	17	10
	2	1	9	6	7	54	44	28	33	10
	7	4	2	8	7	62	45	19	20	28
	4	2	6	3	5	54	54	38	23	9
	5	4	7	9	6	41	36	49	33	13
	1	4	6	8	14	51	53	39	20	19
	5	3	11	6	6	44	46	23	50	9
Mean	3.7	3.6	6.2	6.9	8.9	51.2	45.0	34.8	26.9	13.6
Sigma	1.8	1.3	2.3	2.1	3.4	6.8	8.8	9.6	10.5	5.9

alter this state of affairs, as quite conceivably may have occurred in the case of the eighth-grade subjects.

ITEM DIFFICULTIES

One method of checking the consistency of the stimuli in any test is to analyze the item difficulties of large groups of comparable subjects. If the groups agree in their judgments, the test, even though it be unreliable, may have value in group analysis.

Pitch. Two Stanford groups of 100 each served as subjects for the analysis of this test, as well as for the others. These 200 subjects were chosen by random selection from the Stanford records at hand. The mean score of Group A equalled 45 percentile, and that of Group B, 43 percentile. The correlation between the errors of the two groups was found to be .96 (Pearson r).[26]

[26]Reported in part at the 9th International Congress of Psychology.

TABLE 21
PITCH ERRORS OF GROUP B

Columns	A	B	C	D	E	F	G	H	I	J
Stimulus (d.v.)	30	23	17	12	8	½	1	2	3	5
% errors	2	5	4	10	8	64	51	45	42	14
	1	7	4	12	13	53	53	48	15	7
	3	4	5	11	11	65	27	34	30	11
	4	2	3	12	11	51	42	31	21	15
	1	3	8	4	7	51	56	30	29	10
	4	4	8	5	12	45	46	26	21	21
	3	3	5	5	7	60	50	35	11	15
	6	3	6	8	14	47	39	51	34	16
	3	3	6	13	12	45	51	36	17	13
	2	5	4	6	10	48	37	16	47	12
Mean	2.9	3.9	5.3	8.6	10.5	52.9	45.2	35.2	26.7	13.4
Sigma	1.4	1.3	1.6	3.2	2.3	7.1	8.4	10.0	11.2	3.6

TABLE 22
INTENSITY ERRORS OF GROUP A

Columns	A	B	C	D	E	F	G	H	I	J
Stimulus	5	4	3	2	1	5	4	3	2	1
% errors	3	1	11	9	24	4	0	3	11	10
	1	1	13	10	33	2	1	3	19	23
	3	0	2	21	16	3	3	3	29	25
	2	0	16	18	26	3	5	1	5	30
	0	0	3	18	40	3	2	7	44	10
	1	1	7	13	28	5	2	13	13	28
	0	0	3	5	24	1	4	6	10	50
	0	0	8	16	34	0	5	12	13	42
	2	2	3	5	16	3	7	23	12	23
	0	1	5	13	13	0	3	7	13	37
Mean	1.2	0.6	7.1	12.8	25.4	2.4	3.2	7.8	16.9	27.8
Sigma	1.1	0.7	4.6	5.2	8.3	1.5	1.9	6.3	10.8	12.1

In the main, the stimuli are quantitatively comparable. G3, H6 and 10, I2, and others might be cited as exceptions. Column F has more errors than should occur by chance.

Intensity. The mean score of Group A equalled 54 percentile, and that of Group B, 52 percentile. The

correlation between the errors of the two groups was found to be .94. Just as in the pitch test, the stimuli are fairly well in line. There are certain obvious exceptions, however (I5, etc.). The test is seen to be somewhat too easy for university subjects. More errors were recorded for the second half of the test than for face A.

TABLE 23

INTENSITY ERRORS OF GROUP B

Columns	A	B	C	D	E	F	G	H	I	J
Stimulus	5	4	3	2	1	5	4	3	2	1
% errors	5	1	12	7	31	3	0	7	8	10
	0	0	10	7	32	0	4	2	29	29
	6	3	1	18	15	3	6	6	29	19
	0	1	18	20	27	3	7	2	5	33
	0	1	3	16	42	1	1	2	45	22
	1	0	2	15	38	3	1	4	16	24
	7	0	4	7	25	2	5	4	15	44
	1	1	3	16	27	0	3	7	18	43
	1	2	2	8	23	3	11	17	15	19
	0	1	8	4	16	1	2	8	18	41
Mean	2.1	1.0	6.3	11.8	27.6	1.9	4.0	5.9	19.8	28.4
Sigma	2.6	0.9	5.2	5.4	8.2	1.2	3.2	4.2	11.0	11.0

TABLE 24

TIME ERRORS OF GROUP A

Columns	A	B	C	D	E	F	G	H	I	J
Stimulus	20	20	14	14	9	2	2	5	5	9
% errors	5	10	13	13	26	44	64	29	34	8
	3	0	8	6	16	53	33	28	22	11
	4	3	9	11	16	48	43	27	31	19
	1	4	9	11	14	33	38	30	26	13
	2	8	8	9	25	51	42	25	29	13
	5	2	12	8	27	50	45	28	34	24
	0	2	10	16	21	45	26	33	36	19
	1	6	9	8	25	27	38	33	38	27
	0	1	11	11	25	43	36	30	38	19
	1	1	12	13	20	45	40	33	34	31
Mean	2.2	3.7	10.1	10.6	21.5	43.9	40.5	29.6	32.2	18.4
Sigma	1.8	3.1	1.7	2.8	4.5	7.0	9.4	2.6	4.9	6.9

TABLE 25

TIME ERRORS OF GROUP B

Columns Stimulus	A 20	B 20	C 14	D 14	E 9	F 2	G 2	H 5	I 5	J 9
% errors	4	6	8	14	24	45	56	35	56	20
	1	2	7	11	25	33	30	47	29	16
	8	6	15	9	19	63	54	27	28	17
	1	7	7	11	15	33	40	36	42	15
	3	5	8	7	16	37	38	27	34	17
	11	4	10	11	22	46	58	40	38	23
	4	5	16	10	20	42	32	28	40	17
	6	8	6	9	29	36	47	40	25	15
	3	2	9	7	16	53	33	28	25	14
	6	5	9	11	15	42	47	37	30	20
Mean	4.7	5.0	9.5	10.0	20.1	43.0	43.5	34.5	34.7	17.4
Sigma	2.9	1.8	3.2	2.0	4.5	8.9	9.8	6.5	9.1	2.7

Time. The mean score of Group A equalled 52 percentile, and that of Group B, 47 percentile. The correlation between the errors of the two groups was found to be .89. It is interesting to note that with both groups, Column J was found to be easier than Column E. This may be due to the manner of approach. Before E the progression is from easy to difficult items. Before J the reverse is true.

Tonal Memory. The mean score of Group A equalled 55 percentile, and that of Group B, 46 percentile. The correlation between the errors of the two groups was found to be .97.

Heinlein (28) has made a somewhat similar study. The errors of one of his groups have been tabulated by the present writer, and found to correlate with the errors of Group A to the extent of .93. The mean score of the group in question (Heinlein) equalled 39 percentile. This value is only a rough approximation as is the table which follows (Table 27), since the

writer translated the Heinlein curves into number values. The curves were small as reproduced in the *Journal of Genetic Psychology* so that a magnifying glass was employed. Heinlein has concluded that there are more than five degrees of difficulty in the tonal

TABLE 26
MEMORY ERRORS OF GROUPS A AND B

	Group A					Group B				
Columns	A	B	C	D	E	A	B	C	D	E
Stimulus	2	3	4	5	6	2	3	4	5	6
% errors	4	7	6	45	55	4	10	7	60	74
	4	13	10	31	51	1	16	8	44	46
	7	6	18	39	43	9	19	24	45	36
	4	13	23	28	58	4	25	20	45	76
	1	8	23	32	53	3	9	30	28	57
	2	6	7	48	80	6	9	7	50	81
	4	9	19	33	56	8	14	21	27	62
	2	6	50	23	44	5	11	64	16	57
	6	15	15	41	42	0	20	19	53	46
	2	4	6	18	70	4	2	6	14	77
Mean	3.6	8.7	17.7	33.8	55.2	4.4	13.5	20.6	38.2	61.2
Sigma	1.8	3.3	12.4	9.0	11.5	2.6	6.2	16.5	15.0	14.7

TABLE 27
MEMORY ERRORS IN HEINLEIN'S STUDY

Columns	A	B	C	D	E
stimulus	2	3	4	5	6
% errors	5	11	16	48	66
	13	24	9	28	56
	20	37	12	44	53
	8	23	25	45	66
	8	25	17	32	64
	14	10	15	54	83
	13	33	19	29	82
	14	21	50	37	52
	11	19	24	56	37
	6	14	11	21	78
Mean	11.2	21.7	19.8	39.4	63.7
Sigma	4.3	8.3	11.2	11.2	13.9

memory test. Certain pitch motions are more difficult than others. In other words, the difficulty lies not only in the number of notes, but also in the geometric contour of the melodies which are formed. Rhythmic patterns enter as a factor as well.

Consonance. Means and sigmas were not calculated for the consonance columns since there was obviously no intention of uniform difficulties. The mean score for Group A equalled 39 percentile, and that of Group B, 54 percentile. The correlation between the errors of the two groups was found to be .96. The starred items are the reversals. That is, the majority of ballots were cast for the answer listed as incorrect by Seashore. It must be kept in mind that, in this study, the directions were not followed. A preference test was developed.

On page 314, the writer discussed an attempt to raise the reliability of this test by employing the subjects' ballots as criteria of correctness. However, only those items which showed a value of at least 63% were

TABLE 28

CONSONANCE ERRORS OF GROUPS A AND B

	Group A					Group B				
Columns	A	B	C	D	E	A	B	C	D	E
% errors	1	62*	57—*	1	57—*	0	64*	37	0	50
	6	17	14	16	30	3	9	18	12	17
	5	13	22	36	60—*	6	9	15	8	44
	25	1	3	64*	51—*	27	1	9	67*	51*
	53—*	0	79*	90*	57—*	41	0	75*	97*	49
	20	78*	15	39—	40—	29	69*	12	53*	37
	28	7	61—*	84*	57—*	24	6	63*	82*	58*
	48—	13	63*	56—*	15	42	12	70*	49	15
	4	4	20	16	37	3	9	19	12	20
	43—	52—*	14	22	24	29	57*	7	18	18

retained. The minus lines after certain values of Group A indicate which items were omitted on this account.

A more detailed study along this line has been undertaken by Heinlein (27), who studied the item responses from a musical standpoint. Special attention was paid to interval reversals. The Seashore consonance test is so constructed that each pair has a twin which differs only in the matter of sequence. Thus Pair 1 consists of the major third followed by the major seventh. Pair 40, the twin, is made up of, first, the major seventh, and, secondly, the major third. (Heinlein has pointed out that Pairs 13 and 28 which are reported as reversed in Seashore's book are not, in fact, reversed in the test.) Heinlein's interest, stimulated by an article by Peterson (60), was primarily in finding the effects caused by the reversals. His groups were small, so no comparisons have been made with the writer's data. Heinlein has concluded:

> "From the foregoing analysis, one is inevitably led to conclude that *paired interval comparison is an inadequate method for testing consonance,* since, as evidenced, such comparison is conducive to reaction to the various elements of musical progression, including the feelings which, apparently, are inseparably attached thereto. Assuming that the musically talented have a 'sense' of consonance, and since the object of this test is to determine and point out such musical talent, from the very nature and structure of the test material, there is reason to expect negative results from the talented group."

D. L. Larson (46) has attacked the conclusions of Heinlein. She has charged that the directions were not followed closely enough to keep the preference factor from entering. Her *major* conclusions have

been based on data gathered on 35 musically trained and 35 untrained subjects. In the case of 24 of the stimulus pairs (and presumably this held true with the remaining pairs, although data on these were not given), the trained observers were more consistent in their judgments than were the untrained. Analyses, somewhat similar to Heinlein's, were made on a number of other large groups. Larson has concluded:

> "In the light of correlations which we have found, and in *the light of analysis of intervals in the records of trained*, untrained, and unselected O's, we find that the procedure of paired-interval comparison is adequate for testing consonance. We have not found that the principle of harmonic progression was a sufficiently disturbing factor to distort our score."[20]

It appears to the writer that if this attack on Heinlein is valid, an assumption must be made for the Larson figures. That is, *it must be assumed that the trained* subjects were not basically different from the untrained before formal training. Otherwise, they might conceivably have made more consistent scores (but possibly not as high ones) than they did later when Larson tested them. In other words, although Larson's figures may be indicative, in a critical study one would select and test untrained subjects, train them, and then retest them. It is also conceivable that Larson's findings might be reversed in many individual instances. (The footnote on page 60 of the Larson study has a number of errors which should be corrected to read as italicized: "Pair numbers *13* ($a\ b_b$–$a\ \acute{a}$) and 28

---

[20]C. E. Seashore (71) has replied to certain of Heinlein's criticisms. Seashore is considering a new consonance record, one side of which will test for smoothness, and the other for blending.

(a a′–*a* c′♯) do not appear in the tables, as they are not reversed.")

Heinlein (30) has recently answered Larson. It was impossible, he found, to rule out the feeling of like and dislike in the testing of his subjects. The Seashore directions which ask the listener to judge on a basis of smoothness, blending, and fusion were found to be impossible of enforcement. He pointed out that in this test the procedure of playing a few test items preliminary to the giving of the test is bad, since it really trains the subjects. He emphasized the error of assuming "musical" genes, and the historical fact that consonances change from period to period. The present writer most heartily concurs. Some time ago he stopped giving the test in any form except that of a preference test.

Rhythm. The mean score of Group A equalled 82 percentile, and that of Group B, 75 percentile. The correlation between the errors of the two groups was

TABLE 29

RHYTHM ERRORS OF GROUPS A AND B

	Group A					Group B				
Columns	A	B	C	D	E	A	B	C	D	E
% errors	0	7	11	22	23	5	21	15	12	16
	0	15	6	5	25	7	23	21	15	20
	0	0	7	28	27	4	5	6	25	33
	2	27	27	26	20	5	8	37	12	27
	4	4	48	32	33	11	6	36	45	36
	3	3	6	36	75	6	8	17	39	61
	1	17	6	11	75	27	20	7	13	60
	4	11	8	15	27	10	10	3	17	37
	1	32	27	21	67	4	45	35	35	44
	5	36	22	15	5	8	36	23	19	14
Mean	2.0	15.2	16.8	21.1	37.7	8.7	18.2	20.0	23.2	34.8
Sigma	1.8	12.0	13.3	9.2	23.8	6.5	12.9	12.1	11.6	15.7

TABLE 30

MELODY ERRORS OF GROUPS A AND B

	Group A					Group B				
% errors	5	12	31	7	15	3	11	29	5	13
	9	6	12	34	68—*	7	6	12	32	61*
	25	9	9	65—*	57—*	23	10	8	59*	54*
	4	38—	23	36	62—*	3	40	22	32	58*
	14	27	63—	56—*	42—	11	26	57*	49	45
	8	9	13	12	58—*	8	11	11	11	49
	15	24	25	33	81*	14	24	29	31	73*

TABLE 31

HARMONY ERRORS OF GROUPS A AND B

	Group A					Group B				
% errors	5	7	17	22	40	4	9	17	23	35
	11	3	76*	36	72*	14	2	77*	32	68*
	5	6	41	54*	46	7	5	39—	51—*	43—
	6	6	20	54*	42	11	6	20	47—	43—
	3	9	11	83*	26	6	8	13	78*	30
	53*	6	24	70*	90*	54*—	6	25	68*	86*
	9	60*	46	33	56*	11	51—*	38—	34	53—*

found to be .86. Note that in Column E, rows 6, 7, and 9 of Group A, and rows 6 and 7 of Group B, the errors are more than fifty out of a hundred.

Melody. Means and sigmas were not calculated for the melody columns, since Kwalwasser did not have the items arranged in columns. The mean score of Group A equalled 60 percentile, and that of Group B, 65 percentile. The correlation between the errors of the two groups was found to be .98. The starred items are the reversals, as was true with the consonance test. Similarly, the minus lines indicate the items which were omitted when the attempt was made to raise reliabilities.

Harmony. The mean score of Group A equalled 56 percentile, and that of Group B, 60 percentile. The

correlation between the errors of the two groups was found to be .99. The reversals found in the two Kwalwasser tests indicate that the preferences exhibited by the Stanford subjects fail to check in many instances with the "approved" answers. Of course, a glance at music history will show that tonal preferences are constantly changing from decade to decade.

The writer has made a similar study with Church's fifth-grade data. Correlations between two groups of approximately 90 each were found to be: pitch, .60; intensity, .85; time, .74; memory, .93; consonance, .91; rhythm, .95; melody, .93; harmony, .94.

The mean errors of these fifth-graders were correlated with the mean errors of the writer's Stanford data. This was done to learn whether or not the same errors were made by fifth-graders as by adults. The resulting coefficients were: pitch, .78; intensity, .74; time, .74; memory, .88; consonance, .63; rhythm, .76; melody, .74; harmony, .87. Since the correlation values are fairly high, it may be concluded that there is a general correspondence between the errors made by the two groups.

R. C. *Larson* (47) *has reported* "the percentage right of column scores for the increments of the Seashore pitch test." Grades 5, 6, 7, and 8 were included in this study.

In summarizing, it can be said that the item errors of the Seashore-Kwalwasser battery are quite consistently made. For this reason, the tests are of use whenever group responses to the various stimulus items are to be studied.

Race Differences

Hawaiian Melting Pot. Murdock (54) has arranged a table which showed the percentages of each racial group in Hawaii which overlapped the Anglo-Saxon. The values were: Anglo-Saxon, 50; Anglo-Saxon–Hawaiian, 37; Chinese, 28; Japanese, 28; Portuguese, 10; Korean, 36; Chinese-Hawaiian, 58; Hawaiian, 40; country Japanese, 20.

The present writer has melody and harmony scores on 129 normal-school students from Hawaii. The racial antecedents were quite varied. The melody mean was 26.6 with a sigma of 3.3, as compared with a Stanford white mean of 26.1 with a sigma of 3.2. This gave a D/σ_D of 1.35 (91 chances in 100) in favor of the Hawaiians. The harmonic mean was 22.6 with a sigma of 3.9, as compared with a Stanford mean of 25.4 with a sigma of 2.7. This gave a D/σ_D of 7.2 (100 chances) in favor of the Stanford students.

Negro-White. Kwalwasser (40) has stated that Negroes are superior to white subjects in rhythm and tonal memory, although no figures were given in support of this belief. Kwalwasser possibly drew from the findings of Lenoir (49) on fifth-grade students of Gary, Indiana. One hundred ninety-one white children and 191 Negroes were studied. The latter scored higher on tonal memory, time, and rhythm. By the use of the formula, $D/P.E._D$ values were found of .7, 4.5, and 12.76, respectively (68, 100, and 100 chances in 100). The whites made higher scores in pitch, intensity, and consonance with $D/P.E._D$ values of 3.7,

3.5, and 10.6 (99, 99, and 100 chances). The white subjects averaged 134 months in age and the colored 136.

Davenport (9, 10), in a Jamaica study, has tested farmers of various colors. The pure Whites were largely English and German. To the Whites and Negroes he added a group of mixed bloods, which he called "Browns." The average percentiles on the pitch test were as follows: Black, 26; Brown, 32; White, 17. The Browns ranged from 50 to 100 (raw scores) on the rhythm test. The Blacks averaged 92 percentile and the Whites 69 on this same test.

Peterson and Lanier (62) tested 330 white students from Peabody and 274 Negroes from Fisk on the pitch and consonance tests. $D/P.E._D$ values were in favor of the Negroes in both tests—3.2 (98 chances) in pitch, and 6.8 (100) in consonance. Later tests on 89 Whites and 286 Negroes gave $D/P.E_D$ values of .5 (63) favoring the Negroes in pitch, and 1.1 (77) favoring the Whites in consonance.

These researchers undertook a further study of groups of white students (*N*-372 to 386) from Middle Tennessee State Teachers College at Murfreesboro and Negro students (*N*-288 to 296) from the Agricultural and Industrial Normal College at Nashville. The Whites made higher scores on all the tests with the exception of rhythm, with $D/P.E._D$ values as follows: pitch, 5.84 (100 chances); intensity, 10.10 (100); time, 8.26 (100); consonance, 4.84 (100) tonal memory, 11.80 (100); and rhythm, .12 (53).

Two hundred fifty-eight colored and 219 white

TABLE 32

Test	College students White	College students Negro	Eighth-graders White	Eighth-graders Negro	Fifth-graders White	Fifth-graders Negro
Pitch	81	75	77	73	67	69
Intensity	89	88	83	82	74	75
Time	78	73	71	68	64	64
Rhythm	75	75	70	72	65	66
Memory	74	68	65	54	50	45

students from sixth, seventh, and eighth grades were subjects for Gray and Bingham (25). D/σ_D values were in favor of the white subjects in four of the tests —pitch, 7.1 (100 chances); intensity, 8.1 (100); time, 5.0 (100); memory, 10.3 (100). The Negroes scored higher in consonance —1.2 (88).[27]

G. B. Johnson (33, 34) has tested 3350 Negroes in North Carolina, South Carolina, and Virginia. Medians were given in his data for college students, eighth-, and fifth-graders. In Table 32 Johnson's Negro data are contrasted with the Seashore norms. Johnson did the testing himself. In a footnote he has said:

> "Partly because of the controversial nature of consonance and partly because of lack of time, this test was not given to very many Negro subjects. It is worth stating incidentally, however, that the scores of the 100 Negro college students who did take the consonance test gave a distribution practically identical with the Seashore norms for whites on that test."

He has further stated:

> "It becomes evident that the only fair conclusion to be

[27]Streep (81*b*), in a study of 736 whites and 678 Negroes of the third, fourth, fifth, and sixth grades, found the latter slightly, but consistently the better in the rhythm and consonance tests. As she herself says, there is the possibility that the Negroes were sufficiently older to account for the differences.

TABLE 33

	Group 1	Group 2	Group 3
		Pitch	
Mean	76.9	70.2	81.7
σ	12.6	10.5	7.5
N	36	53	200
D/σ_D	1 and 2, 2.7 (100 chances)	1 and 3, 2.3 (99)	
		Consonance	
Mean	33.6	29.6	34.5
σ	3.5	3.7	3.3
N	36	53	200
D/σ_D	1 and 2, 5.0 (100)	1 and 3, 1.5 (93)	
		Melody	
Mean	27.0	23.7	26.1
σ	3.3	3.8	3.2
N	36	54	200
D/σ_D	1 and 2, 4.7 (100)	1 and 3, 1.5 (93)	
		Harmony	
Mean	21.3	18.7	25.4
σ	3.6	3.6	2.7
N	36	54	200
D/σ_D	1 and 2, 3.2 (100)	1 and 3, 6.8 (100)	

drawn from the data is that there are no significant differences between whites and Negroes in those basic musical sensibilities measured by the Seashore tests."

Peacock (59) has concluded that Whites surpass Negroes in Seashore test ability.

American Indian-White. Garth and Isbell (18) tested 409 mixed-blood and 360 fullblood American Indian children. The experimenters concluded that, taking other factors into account, there existed no real racial differences.[28]

Chinese and Japanese-White. The present writer has tested 36 Japanese and Chinese students of the University of California and Stanford who were born and

[28]Garth's data can now be found in (17*a*).

reared in America (called Group 1 in Table 33) and contrasted them with 53 others of similar nationalities and scholastic status who have lived in America for considerably shorter periods of time (Group 2). In Tables 33 and 34, the Stanford white students make up Group 3.

It is evident that the Stanford white subjects scored higher on all but the melody test. However, their mean score of 34.5 on the consonance test is higher only because one half the subjects were girls. The white men and the members of Group 1 (all were male) made exactly similar scores. The members of all three groups had been subjected to "intelligence" tests a number of times. They understood English perfectly. The present writer tested every subject personally, and felt that he had had excellent cooperation. Groups 1 and 2, however, may not have had quite the motivation of Group 3, since the members of the former two were not in the writer's classes. While they were being tested in fraternity houses, they seemed extremely interested.

The slight alteration in motivation may conceivably account for the differences between Groups 1 and 3, but not between 1 and 2. One might guess that varying acquaintance with occidental music may have made at least a portion of the difference. Such a hypothesis would seem to do quite well for preference tests such as consonance, melody, and harmony, but does not seem to fit what is known of pitch. Still, ethnological evidence from China indicates that pitch differences are not stressed as they are in the Occident.

TABLE 34

	Chinese		Stanford Whites
		Pitch	
Mean	64.1		81.7
σ	9.8		7.5
N	43		200
D/σ_D		11.0 (100 chances)	
		Consonance	
Mean	26.0		34.5
σ	3.3		3.3
N	44		200
D/σ_D		17.0 (100)	
		Melody	
Mean	23.5		26.1
σ	3.4		3.2
N	44		200
D/σ_D		4.3 (100)	
		Harmony	
Mean	17.7		25.4
σ	3.5		2.7
N	44		200
D/σ_D		15.4 (100)	

This apparent fact may have a bearing on the problem.

A small group of Chinese college students were tested at Hong Kong. The data are given in Table 34.

The writer would hazard this conclusion from his work on oriental students. There is nothing to indicate that oriental blood per se would predispose one to lower scores on the tests in question. It is conceivable, however, that lack of acquaintance with occidental music may cause a lowering of scores.

The data on the other racial groups are obviously too confused to form bases for generalizations. The factors of test unreliability and unequal motivation are probably responsible for much of the lack of agreement.

Training and Validity

Training—Theory and Experiments

The Term "Training." The term "training" has been employed in a variety of ways. (*a*) Certain writers refer to the giving of a music test a number of times as "training." (*b*) In addition to the retesting, the subject may be told which of his responses are correct and which are incorrect. (*c*) "Training" may include explanations of the concepts involved, cues as to procedures, and the like. (*d*) It may mean simply the accidental stimulation of the organism over long periods of time, with no formal instruction. For example, the Arabs have, in the main, heard less jazz than have Americans. (*e*) "Training" may refer to the formal musical training of an individual. Obviously, the efficacy of the training will vary with the variety employed.

Capacity versus Ability. The question of training leads to a discussion of the concepts of capacity and ability. As defined by Seashore (67): "The term 'capacity' has reference to the inborn or native power; the term 'ability' is used to designate acquired skill in the use of a capacity." It should be obvious, then, that the usual aim of the tester is to measure capacity. He should hesitate to label his findings "measures of capacity," however, as he rarely, if ever, knows the exact extent to which acquired skill has functioned. When he states that one's musical ear cannot be made keener, he is usually referring to the capacities of pitch, intensity, etc. He may or may not be correct.

The musician well knows that, after formal musical training, many students can more easily tell when they are on the key. This may be merely a shift from cognitive to physiological limit, or it may mean the entry of some other factor.

Training Experiments. For the conclusions in his book regarding the effect of training on pitch scores, Seashore (67) has depended very largely on several studies which were carried on prior to and so without the use of his phonograph record. (*a*) Buffum (67)

> "classified 28 eighth-grade pupils in a preliminary fifteen-minute individual test for each one and then proceeded to train them by requiring 40 minutes of the most specific and intensive practice under favorable conditions for 20 successive days, recording the average achievement for each day. The result was greatly surprising. Instead of showing progressive improvement with practice, each child remained as rated in the preliminary test, with only two exceptions, and the average ability for the class was the same at the end of the 20 days of this intensive practice as at the beginning."

(*b*) F. O. Smith (74) has done a somewhat similar piece of work with 476 children, which corroborates Buffum. He has concluded:

> "The sensitiveness of the ear to pitch difference cannot be improved appreciably by practice. There is no evidence of any improvement in sensitiveness to pitch as a result of practice."

Whipple, in an editor's note attached to one of the present writer's articles, has criticized these results as follows:

> "The interpretation of results in such mass experiments as those of Buffum and Smith is, in my judgment, decid-

> edly difficult, if not often misleading. Certainly, an equally important method of studying the effects of practice is to confine one's effort to drilling a competent, though unmusical, adult under laboratory conditions which permit some measure of qualitative analysis of what takes place. If this is done, as I myself showed years ago in a study of pitch discrimination (*Amer. Jour. of Psych.*, 14: 1903, 561-568) not listed in the bibliography here, the effect of systematic training of an unmusical person is a curiously complex one—there appears marked improvement under certain conditions, along with none at all under others. Without going into all the details, suffice to say that I was able to train my subject to discriminate correctly pitch differences of less than three vibrations (a perfectly laudable achievement), but could at will alter conditions slightly and cause her to fail utterly to discriminate pitch differences, not of a few *vibrations*, but of *more than two octaves*. What is one to conclude from that? Can the unmusical be trained or can they not?"(12).

Without question, there is much to favor a qualitative study over one which employs group results.

Ross (63) found that the discrimination of time, as measured in the Seashore test (old form, not the Columbia record), displayed a marked improvement in the case of eight subjects when the test was repeated seven times on successive days. In a somewhat similar study, Wright (88) has ascertained that the giving of the Seashore battery of tests daily for a week increased the average scores slightly, but not to any significant extent. In a number of cases the highest score was made in the middle of the week.

According to the work of Vance (84), pitch discrimination is most keen at 128, 256, and 512 d.v.—just the octaves most used in music. In fact, the musi-

cally trained subjects had better "musical ears" at these pitches than had the non-musical, but were poorer at higher and lower pitches.

Seashore (67) has stated that training undoubtedly favors improvement in the discrimination of consonance, but that the improvement is not so great as is commonly supposed. He has further cited an experiment in which 14 observers were subjected to a prolonged series of specific and intensive practice in discriminating intensity differences. There were 14 to 30 periods of one-half hour each. On the whole, the achievement of these 14 observers was practically the same at the end of the practice periods as at the beginning. In another study he experimented with 15 students who had been blind from infancy. These subjects were compared with 15 "seeing" students in pitch and intensity discrimination. The average scores were approximately the same. Apparently the intensified training which the blind had had since childhood (assuming them to be typical) had not tended on the whole to make them keener in these abilities.

Seashore and Mount (72) have reported correlations between musical, instrumental, and vocal training (as judged by questionnaires) and the Seashore battery which range from .39 down to .03. Correlations between musical environment (as judged by questionnaires) and the battery ranged from .30 to —.03.

Travis and Davis (83) have found that "the *sense of pitch*, the *sense of intensity*, and *tonal memory* enter into the function of speech."

In a study by Brennan (4), training was found to

correlate significantly (.40 or more) with pitch and rhythm discrimination, and with tonal memory. The training score was made to be the number of half-hour lessons the student had had. It is well to bear in mind that causality is not necessarily expressed in such a correlation relationship. As Brennan has said: "In other words, his tonal memory need not be good because he has received a certain amount of training, but because he is possessed of a keen tonal memory, he naturally gravitates toward thinking and working with tones." The present writer would amend the last line to read "he *may* naturally gravitate, etc." Brennan also found Spearman rank correlations ranging from $.17 \pm .14$ to $.47 \pm .12$ between Seashore scores and music performance as rated by expert judges.

Using a similar training score criterion, De Graff (11) found correlations between it and rhythm discrimination of .09 with 464 adults, of .10 with 282 eighth-grade pupils, and of .21 with 272 fifth-grade pupils.

Motivated by a desire to improve on Brennan's controls, Brown (7) ran correlations between teachers' judgments of musical ability and scores on the Seashore battery. They ranged from $.11 \pm .06$ to $.41 \pm .05$. The teachers' judgments and average score on the test battery correlated at $.38 \pm .05$ (reduced to .35 when age and intelligence were partialled out).

Gray and Bingham (25) called their subjects musically trained if they had had two lessons a week for a year or more. Their trained groups scored slightly higher in all of the Seashore tests with the exception

of intensity and time where occasionally the untrained made higher scores.

Malmberg (50) carried on work which eventually led to the construction of the Seashore consonance test. In that preliminary study he found a Spearman *R* of .02 between the test score and amount of training (judged by questionnaires), and an *R* of .06 between the test score and ability to perform (also judged by questionnaires).

As has been reported earlier, D. L. Larson (46) found that, in the main, her musically trained subjects scored above the untrained on the consonance test.

Klauer (37) put subjects through two months of intensive marching, clapping, beating time, etc., and failed to find a significant increase in scores on rhythm discrimination over scores made by untrained control subjects. However, her results are complicated in that her control group was retested in a much shorter time than the training group and was obviously atypical. That is, its mean percentile was considerably lower than that of her training group so that no accurate comparisons can be made.

R. H. Seashore (73) has reported a correlation of $.82 \pm .02$ between scores on the rhythm test and learning rhythm. The subjects in this experiment comprised those who stood in the highest and the lowest 10 per cent on the rhythm test.

In a study somewhat similar to that of Klauer, Baugh (2) used two very small groups of children who ranged in age from 7 to 10, and in IQ from 78 to 118. The children tapped out rhythm, were told of

the nature of rhythm, worked somewhat with various times, etc. While in the retest, the control group lost four points, on the average, the work group gained six points.

Validation

Critical Discussion of Validation. Garrett (17) has said that the "validity" of a test "depends on the *fidelity* with which it measures whatever it purports to measure." The difficulty in ascertaining validity arises in the quest for some independent criterion of the phenomenon in question. For example, when is a test valid in probing for time limens? The Seashore test was designed to measure the ability to detect differences in empty time. Would a test in which the tonal stimuli are held for varying lengths of time give the same classification of subjects? No one knows. In fact many types of tests could be constructed, and be logically called time tests. But which would be the best time test?

A similar difficulty is met with in "intelligence" testing. There are many tests which purport to measure "intelligence," but no tangible criterion is available. After all, psychological tests are being constructed for practical situations—college entrance, clinical classification, etc. It thus appears to the writer that, for the time being at least, validation must rest on work such as Stanton is doing at Rochester. The validation, then, would not be for time, but for success in musical conservatories (see page 372).

Data on Validation. In a study of 24 music students, Wright (88) found correlations of .45, .51, and .73

with achievement tests (largely dictation, and piano performance) devised by herself. A sight-singing test devised by Gaw (23) was found to correlate .46 with pitch, .36 with intensity, and .56 with memory (N=149). Mosher (53) ran correlations between music achievement scores and the Seashore tests on about 450 students. His values were: time, .36; consonance, .29; pitch, .44; memory, .44; and intensity, .49. Tierney (82) also correlated the battery with an achievement test of her own construction. The value was found to be .22±.06.

Beach (3) has reported that when the Seashore tests were employed in the St. Louis schools "the accelerated groups were much superior and progressed more rapidly."

Barnes (1) claimed to have found "a high correlation" between the Seashore battery and grades in music courses in a southern college.

Heinlein (29) has found that the pitch record and his chordal discrimination test measure different behavior variables (.10±.12). In a study of the seventh grade, R. C. Larson (47) has correlated the Kwalwasser-Ruch with the Seashore. When inverse *S.D.*'s were used the value was .55; with T-scores it was .54; with regression weights .73.[20]

[20]McCarthy (50*a*), with a population of 71 children, found correlations as follows with grades in singing: pitch, .43; intensity, .33; memory, .40; and consonance, .41. With a population of 58, correlations were run against results on a music questionnaire. The coefficients were: pitch, .57; intensity, .31; and memory, .83. In the same study she found that pitch and intensity scores showed no improvement after four repetitions. The memory test scores, however, showed a marked increase.

Hammond (26*a*) ran correlations between the rhythm measure and Test 6, Sense of Rhythm, of the Logasa-McCoy-Wright Appreciation of Literature Test. The value found was .18±.11.

In a recent study, Salisbury and Smith (65) devised a test of sight-singing ability which has high reliability. They were also interested in the Seashore battery as prognostic of sight-singing ability.

> "Pitch and tonal memory were found to have very significant predictive value, the correlations with the criterion being between .60 and .65, respectively, and in the regression equations involving other prognostic materials, took significant weights. The other four Seashore tests were of *decidedly* less value for the purpose at hand."

The abilities to sing on key, to sing intervals, and "voluntary control" were found by Brennan (5) to correlate somewhat with pitch and tonal memory ($.21 \pm .09$ to $.38 \pm .05$).

Highsmith (31) has concluded from a study of college music scores:

> "2. The validity of the tests when measured by school marks is very low, .312. 3. The intelligence tests used (Terman Group and Thurstone) gave a better prediction, r equals .423, of probable success in music than the Seashore tests. . . . . 5. On the whole the Seashore tests proved of doubtful value as means of predicting the probable success of students in music."

The validity of the Seashore tests has been studied since the academic year 1923-1924 at the Eastman School of Music, University of Rochester. This work has been under the guidance of Stanton who has published a number of monographs concerning her work (76, 77, 78, 79, 80). Stanton's methods have been followed by Church (8) with apparent success.

One of Stanton's first studies displayed the unreliability of music grades as criteria of musical achievement. This unreliability she was able to remedy to

some extent. After considerable work in trying to predict success in the music courses, it was found best to combine the Seashore battery with the Iowa Comprehension Test—a college "intelligence" test. By this means a series of critical score combinations was formed, below which the academic mortality was so high that it seemed advisable to exclude in the future all applicants with scores below these critical values. This procedure is a standard method in many colleges. It has been found quite efficacious in keeping out students who would probably have failed had they been enrolled. From a *practical* standpoint it works but little injustice. It presupposes, of course, that the finances of the school in question are limited.[80] However, the writer would like to mention in passing that certain of his friends have made at least fair musicians of subjects who scored in the very lowest percentiles of the Seashore pitch test. It is only fair to state that relatively large amounts of time and ingenuity had to be expended. Conservatories cannot afford to work with one subject when another is at hand who gives promise of progressing at a far greater rate. Yet all musical opportunity should *not* be kept from the low scorer.

Conclusions. In any training study the procedures should be carefully outlined. Tests are always made on trained or semi-trained subjects and so reveal abil-

[80]Stanton and Koerth (81*a*) have found the coefficient of mean square contingency for T_1 and T_2 composite scores (retests after three years of intensive conservatory training) in terms of the Eastman profile classification for 157 students. The value was .695 out of a possible .866.

ities. The extent to which they indicate capacities is not known. Retesting and the methods so far employed in short "training" periods do not seem, in the main, to change test scores appreciably. Nothing definite is known concerning the results of long exposure to a musical environment. To show the effects of *this or of formal musical training, tests must be given* before and after the musical exposure. Since grades in music courses are unreliable, it is inadvisable to employ them as criteria of musical achievement. The procedure (adopted by the Eastman School of Music) of combining the Seashore battery with a standard "intelligence" test, and then excluding all applicants who score below a value previously determined as one which only failures make, seems to be a *practical* plan, and to work but little injustice when time and money are limited. In view of the difficulties which arise in the validation of tests when names are given which stand for intangible phenomena, Stanton's work should be considered as a start toward validation.[31 32 33]

---

[31]In a recent University of Iowa study, W. S. Larson (48) has said: "Experimentation is now under way in a public school system for the purpose of investigating the practical application of these aptitude tests (Seashore tests) as a selective agency. We hope that results may be obtained that will be basic not only for the classification of students in instrumental music, but will differentiate talents for other aspects of school music as well."

[32]In October, 1928, M. E. Wilson (86) gave the Seashore tests to music students of Ohio State University. These test scores were correlated with "averages of grades from several courses taken simultaneously or during the three quarters of the school year." The averages of the Seashore tests correlated $.21 \pm .08$ with applied music, $.42 \pm .06$ with ear training, $.21 \pm .09$ with theory, $-.02 \pm .11$ with teaching courses, $.26 \pm .09$ with appreciation, and $.25 \pm .07$ with the total average of grades (86).

Elemental Character of Tests

The difficult problem of the relation of nature to nurture has been mentioned in the section on "Capacity versus Ability."

It is a truism that a new life-history begins with each fertilized ovum. The ovum is a mass of chemicals. These chemicals interact with each other and with forces outside the cell walls, but have not in themselves *abilities,* as much loose talk might indicate. Obviously, the geneticists, in their studies of human capacities, can only hold the environment as constant as possible or else vary it in prescribed ways, note the characteristics of parents, and then watch for similar behaviors or structural resemblances in the offspring. In the present condition of the social sciences the methods of environmental analysis are inadequate to measure the forces which have operated on a child up to the fifth school year (when the Seashore tests can first be given). Thus a resemblance (for example, a score on the Seashore pitch test) in parent and offspring does not necessarily mean that the chemical natures of the germinal cells were similar or were predisposed to react with the "normal" environmental forces in such a manner as to make the test scores in question. Home, school, and church, the influences of friends, and countless other less obvious factors may have operated to boost or lower score values. Economic motives are

[33]Just as this monograph went to press a study by R. C. Larson (47) appeared. In it will be found two studies pertinent to this section: (*a*) data on changes in Seashore score after training in a music school; (*b*) important data on validation.

frequently operative, as well. So a study like Stanton's, reported in the next paragraph, cannot be conclusive.

Stanton (76), aided by an appropriation granted by the Carnegie Institution of Washington, studied the abilities of 85 parents and offspring on four of the Seashore battery—pitch, intensity, time, and tonal memory. "These measures were supplemented by a systematic interrogation which covered questions in regard to musical environment, musical education and training, musical activity, musical appreciation, musical memory and imagination." Her results are stated as follows: "The harmony of the results with certain Mendelian laws in the family distribution tables of assumed gametic formulae is not improbable."

Certain work was done by Smith (74) on pitch discrimination prior to the construction of the phonograph record. Tests were given to two large groups of siblings. One group gave a correlation of .48±.03; the other group gave a value of .43±.04. In each case a younger child was paired with an older sib. When the younger children of Group 2 were paired with other children of the same age and sex as their sibs, but who were not related to them, the correlation was .53±.03. These correlation values may possibly be explained in part by errors in selection. It is quite likely that, on the average, the younger the child the younger the sib with whom he was paired. Since age is an important factor in test scores of this type (see the Seashore test age-norms), this would in itself give spurious positive correlations. The reliability of this test is unknown.

According to Seashore and Mount (72), "to the extent that the test is elemental, there should be no appreciable improvement with practice, development with age or variation with general intelligence." It has been seen that there is practically no relationship between results on "general intelligence" tests and any of the Seashore-Kwalwasser battery. The discussion of practice did not give a very concise answer. Test scores do improve with age. Of course, it may be true, as Seashore has claimed, that the physiological limits do not improve. There is no obvious method of checking this, however.

Modifications of Test Battery

The following attempts at modification of the tests in question have been discussed previously: by Gaw, by McGinnis, by Poorman, by Salisbury-Smith, and by the writer in his Stanford study.

Gaw (23) has found that the Seashore time and tonal memory tests are considerably too difficult for fifth-graders. She has modified the time test by extracting Columns A, B, C, D, E, and J from the record and playing them twice. The resulting score curve was found to be more nearly normal. Certain percentiles have been found for 53 children. The tonal memory test has been modified by the rendition of Columns A, B, and C three times. Percentiles have been figured for 141 fifth- and sixth-grade children.

The point has been previously made that subjects are very likely to judge consonance differences on the basis of affective tone and not from the standpoint of

consonance (as Seashore defines it). The responses "better, worse" favor this error. D. L. Larson (46) has therefore suggested the responses "less, more."

The writer has found that numerous students become slightly confused at the Seashore directions, "whether the second is higher (weaker, etc.) . . . than the first." It seems much simpler to say, "If the pitch of the tonal sequence goes higher, mark *h*, etc." In giving small children the tonal memory test confusion is sometimes avoided by offering to each child five slips of paper numbered one through five. The subjects are asked to point to each slip in turn as the tones are played. This method apparently allows them to attend more to the tones and less to the act of counting. If these changes in method alter the norms, the writer has no record of it.

Highsmith's (31) modifications of the tests were also previously discussed. By throwing together results from all the Seashore tests, the two halves, when stepped up by the Spearman-Brown formula, gave a reliability of $.89 \pm .05$.

The writer has given the intensity test on the Brunswick Panatrope at full intensity. One hundred eight members of a class in elementary psychology served as subjects. They had not taken the test previously. The median approximated 87 percentile, and the mean, 83. A comparison with the intensity mean data which were reported earlier showed that the D/σ_D equalled 10.57. This indicated that the more modern instrument altered the standardization enormously. There was probably no selective error in the type of

subjects, since heretofore, in similar classes, the mean scores had been quite constant.

With 249 Mills College entrants as subjects, Gaw (23) gave the pitch test on the old type Brunswick. The test was immediately repeated on the Panatrope, with a correlation of $.732 \pm .02$. The Brunswick mean approximated 57 percentile and the Panatrope 67. The $P.E_D$ of means was .405 (62 chances). However, as has been previously suggested, the repetition at the same sitting may have introduced other variables.

In an attempt to get the rhythm median lowered somewhat, the writer changed the phonograph speed from 78 r.p.m. to 60. The attempt failed, as the median and mean approximated 77 and 79 percentiles. This work was on a group which had not been previously tested. By comparing the mean with that reported earlier on a similar group, it was seen that the new one had only 74 chances in 100 of being larger than the old one.

Both Kwalwasser (39) and Schoen (66) have added tests of their own construction to make up the supposed incompleteness of the Seashore battery.

Critical Evaluation of the Tests

Pitch. This test has a reasonably high reliability. It has certain elements in common with the test of tonal memory, but little in common with the other tests of the battery or with "intelligence" tests. Its standardization appears good. Gilliland and Jensen (24) have carried on a detailed study of the test, and have made this declaration:

> "The Seashore records have been standardized and are no doubt very satisfactory for the purpose for which they are used in school room tests. . . . But for laboratory experiments on pitch differences the phonograph records are not satisfactory."

A number of what might at first glance be called "questionable features" of the test can be noted.

1. The intensity variable is not perfectly controlled.

2. In the present scoring system each item is given an equal weight. It might be possible to construct a weighting system which would improve reliabilities. However, the writer is inclined to feel that such a change would probably not be worth while.

3. A subject's score does not always indicate his true ability, as is shown by the frequent change of score when the test is repeated. For clinical purposes this difficulty can be partially overcome by repeating the test several times (on different days). If the score fluctuation is great, the median of a number of trials *may possibly* be nearest the true ability score. If the improvement is marked the scores on the last test *may* have most significance, but may show learning or recollection of prior answers to some extent. Indeed, the writer knows of no absolute method of determining the true score in a clinical case. However, such refinement of measurement is probably not very essential; since about all one usually cares to know is the quartile in which the subject's score falls.

4. It is occasionally obvious that the possessor of absolute pitch need hear only the first note of each series of two tones to figure the answer. The writer can see no harm in this nor defect in the test, since such

a phenomenon merely indicates that the subject has great accuracy of pitch discrimination, and sometimes a memory for absolute pitch.

5. Music teachers are prone to say that the differences displayed by the test are psychological rather than musical (55, and others). To a certain extent their claim is justified, since practical music does not often make use of the minutest differences registered on this test.

6. The work of Lanier (43) might conceivably indicate that a decrease in the length of the test by at least one-half would entail no loss in reliability. This would greatly aid in relieving the monotony. However, the writer has already questioned Lanier's criterion of reliability.

7. The test elements are simple auditory stimuli, not snatches of music. This fact should really be a point in the test's favor if it is to be a "capacity" test.

8. According to R. C. Larson (47), "the results show that the 5 d.v. column functions little if any in the test, and that there is a need for a finer choice of increments between the 2 and 8 d.v. range."

Seashore (67, 68) has given the following advice concerning the pitch limen disclosed by his test.

> "Below 3 d.v.: May become a musician. 3-8 d.v.: Should have a plain musical education. 9-17 d.v.: Should have a plain education in music only if special inclination for some kind of music is shown. 18 d.v. and above: Should have nothing to do with music." Another way of stating approximately the same thing is as follows: "Best 10 per cent: stimulate enthusiastically. Next 20 per cent: encourage freely. Next 40 per cent: encourage. Next 20 per cent: question. Next 10 per cent: discourage."

This advice is probably the best that can be given today, if time and money are limited.

Intensity. The majority of the statements made concerning the pitch test are equally applicable when intensity is considered. The test is, however, considerably less reliable.

Time. The reliability of this test is quite low. Suggestive effects are quite likely to enter in giving the time record as a group test, since subjects are prone to nod heads, or beat out the times on desk-tops. This test is the only one given at a speed other than 78. There is sometimes a difficulty in making the directions clear. For fifth-graders this can frequently be facilitated by moving the hand at a constant rate over the blackboard and by marking with a crayon at different time intervals. This gives a visual as well as an auditory picture. The seeming unmusical manner of judging time is decidedly disturbing to many students who have had musical training.

Consonance. Compilations of data concerning this test are of questionable value since various subjects react to its directions in such different manners. Even D. L. Larson's (46) reliability figures are low. Heinlein (27, 30) and Guernsey (26) have discussed the error of arbitrarily fixing consonance values. In view of the complicated directions which are historically in error even when they can be understood, the writer has treated the record as a straight preference test. The intensity factor is not constant. This predisposes certain stimuli to lower or higher ratings than they would normally receive (15). A few of the combinations

appear on occasion to be so hideous that laughter is apt to ensue. The noise thus engendered is likely to deaden the following stimulus pair. However, for all its faults and unreliability, it is well to keep in mind that the errors made by groups of subjects are quite consistent.

Tonal Memory. Tonal memory shares with pitch the honor of being the most reliable member of the Seashore-Kwalwasser battery. As has been mentioned before, the two tests have certain elements in common. From one viewpoint, all the tests are "memory tests." When giving the tonal memory record to children, certain difficulties occasionally arise. Children are likely to count the tones aloud, or in stage whispers, as they are given. This difficulty can be at least partially eliminated by giving them numbered papers to manipulate. In this test the subject is informed (indirectly) of the number of notes to be anticipated in the succeeding columns. As the number gradually builds up—2, 3, 4, 5, and finally 6—the murmurs of discouragement are frequently audible. Heinlein (28) concludes as follows:

> "2. The method for grading the test is inadequate.
> 3. Tonal memory of the identification type has little value in the diagnosis of musical talent."

Heinlein's statement may be correct. However, the writer feels that this test has at least psychological significance.

Rhythm. The rhythm test is the newest member of the Seashore battery and the poorest in accuracy of standardization. Its reliability is very low. One

should be very cautious, therefore, in employing the test. For educational ratings of this and the other Seashore tests, the readers are referred to Kelley (36, p. 276).

Kwalwasser Battery. The Kwalwasser tests in their present form seem too unreliable for purposes of clinical diagnosis. For an adverse critical evaluation of this and the Seashore battery, see Norris (57).[34] For a statement of music-testing aims, see reference (56).

General Conclusions Concerning the Use of the Tests. The Seashore pitch and memory tests would appear to possess sufficient reliability for certain diagnostic purposes. The Kwalwasser battery and the remainder of the Seashore battery should be employed only with extreme caution. The item difficulties of all members of both batteries are highly consistent. For certain group studies they should, therefore, prove of value. For purposes of classification the use of a battery composed of the entire group of tests would also appear to be justified, at least from the standpoint of reliability. Kelley (35) has given a formula (number 147) for obtaining the reliability of a battery if the reliabilities and the intercorrelations of the several constituents are known. If the following reliabilities—pitch, .75; intensity, .66; time, .51; consonance, .65; rhythm, .47; and memory, .83—and the writer's table of intercorrelations are assumed (the reliabilities to be later raised by the Spearman-Brown formula), and it is further granted that the weights are taken inversely

[34]A more recent article which is also somewhat opposed to the Seashore-Kwalwasser battery is by J. C. Moos (52).

as the sigmas, then r equals .885. If the two Kwalwasser tests are added, with assumed reliabilities of .42 and .21, the battery reliability becomes .870.

References

1. Barnes, J. C. Study reported by J. Peterson in Proceedings of "The Twenty-fourth Meeting of the Southern Society for Philosophy and Psychology." *Amer. J. Psychol.*, 1929, 41, 507.
2. Baugh, E. Development of rhythm perception through training. Thesis, Ohio State Univ., 1928.
3. Beach, F. A. Some comparative measures of intelligence, musical capacity and achievement. *Proc. Music Teach. Nat. Asso.*, 1924, 112.
4. Brennan, F. M. The relation between musical capacity and performance. *Psychol. Monog.*, 1926, 36, No. 167, 190-248.
5. ———. A report of three singing tests given on the tonoscope. *Psychol. Monog.*, 1926, 36, No. 167, 249-262.
6. Broom, M. E. A note concerning the Seashore measures of musical talent. *School & Soc.*, 1929, 30, 274-275.
7. Brown, A. W. The reliability and validity of the Seashore tests of musical talent. *J. Appl. Psychol.*, 1928, 12, 468-476.
8. Church, E. Use of tests and measurements in grouping music students of the junior high school. *Music Supervisors J.*, 1929, 16, 79.
9. Davenport, C. B. Race crossing in Jamaica. *Scient. Mo.*, 1928, 28, 225-238.
10. Davenport, C. B., & Steggerda, M. Race crossing in Jamaica. *Publs. Carnegie Instit., Wash.*, 1929, 5.
11. De Graff, L. H. Norms of the sensitiveness to rhythm. Thesis, Univ. Iowa, 1924.
12. Farnsworth, P. R. The effects of nature and nurture on musicality. *27th Yrbk. Nat. Soc. Stud. Educ.*, 1928, Pt. 2, 233-245.
13. ———. The Spearman-Brown prophecy formula and the Seashore tests. *J. Educ. Psychol.*, 1928, 19, 586-588.
14. ———. Reply to Dr. Lanier's note on the Seashore tests. *J. Educ. Psychol.*, 1929, 20, 693-694.
15. Farnsworth, P. R., & Voegelin, C. F. Dyad preferences at different intensities. *J. Appl. Psychol.*, 1928, 12, 148-151.
16. Fracker, G. C., & Howard, V. M. Correlation between intelligence and musical talent among university students. *Psychol. Monog.*, 1928, 39, No. 178, 157-161.

17. Garrett, H. E. Statistics in psychology and education. New York: Longmans, Green, 1926. Pp. 330.

17a. Garth, T. R. Race psychology. New York: Whittlesey House (McGraw-Hill), 1931. Pp. xiv+260.

18. Garth, T. R., & Isbell, S. R. The musical talent of Indians. *Music Supervisors J.*, 1929, 15, 83, 85-87.

19. Gaw, E. A. A revision of the consonance tests. *Psychol. Monog.*, 1918, 25, No. 108, 134-147.

20. ———. A survey of musical talent in school of music. *Psychol. Monog.*, 1922, 31, No. 140, 128-156.

21. ———. Some individual difficulties in the study of music. *J. Educ. Res.*, 1922, 5, 381-388.

22. ———. Individual differences in musical sensitivity. *Bull. Mills Coll.*, 1925. Pp. 52.

23. ———. Five studies of the music tests. *Psychol. Monog.*, 1928, 39, No. 178, 145-156.

24. Gilliland, A. R., & Jensen, C. R. The reliability of the Seashore phonograph record for the measurement of pitch discrimination. *J. Exper. Psychol.*, 1922, 5, 214-222.

25. Gray, C. T., & Bingham, C. W. Musical ability of colored and white pupils. *J. Educ. Psychol.*, 1929, 20, 501-506.

25a. Griffitts, C. H. Fundamentals of vocational psychology. New York: Macmillan, 1924. Pp. 372.

26. Guernsey, M. The rôle of consonance and dissonance in music. *Amer. J. Psychol.*, 1928, 40, 173-204.

26a. Hammond, M. A note concerning rhythm tests in poetry and in music. *J. Appl. Psychol.*, 1931, 15, 90-91.

27. Heinlein, C. P. An experimental study of the Seashore consonance test. *J. Exper. Psychol.*, 1925, 8, 408-433.

28. ———. A brief discussion of the nature and function of melodic configuration in tonal memory, with critical reference to the Seashore tonal memory test. *J. Genet. Psychol.*, 1928, 35, 45-61.

29. ———. The affective characters of the major and minor modes in music. *J. Comp. Psychol.*, 1928, 8, 101-142.

30. ———. Critique of the Seashore consonance test: a reply to Dr. Larson. *Psychol. Rev.*, 1929, 36, 524-542.

31. Highsmith, J. A. Selecting musical talent. *J. Appl. Psychol.*, 1929, 13, 486-493.

32. Hollingworth, L. S. Musical sensitivity of children who test above 135 I.Q. *J. Educ. Psychol.*, 1926, 17, 95-109.

33. JOHNSON, G. B. A study of the musical talent of the American Negro. Thesis, Univ. N. Car., 1927.
34. ———. Musical talent and the Negro. *Musical Supervisors J.*, 1928, 15, 81, 83, 96.
35. KELLEY, T. L. Statistical method. New York: Macmillan, 1923. Pp. xi+390.
36. ———. Interpretation of educational measurements. Yonkers, N. Y.: World Book Co., 1927. Pp. xiii+363.
37. KLAUER, N. Effect of training in rhythm upon rhythmic discrimination in the intermediate grades. Thesis, Univ. Iowa, 1924.
38. KUHNES, D. M. An experimental study of musical talent. (*Bull.* No. 2) Akron, Ohio: Teach. Coll., Univ. Akron, 1928.
39. KWALWASSER, J. Manual of directions for Kwalwasser tests of melodic sensitivity and harmonic sensitivity. Syracuse, N. Y.: Author (Univ. Syracuse), 1926. Pp. 3.
40. ———. Tests and measurements in music. Boston: Birchard, 1927. Pp. 159.
41. ———. Tests foretell your probable success or failure in music. *Musician*, 1927 (July).
42. ———. Tests and measurements in music. *Psychol. Bull.*, 1928, 25, 284-301.
43. LAINIER, L. H. Prediction of the reliability of mental tests and tests of special abilities. *J. Exper. Psychol.*, 1927, 10, 69-113.
44. ———. Note on reliability predictions for the Seashore music tests. *J. Educ. Psychol.*, 1929, 20, 691-692.
45. ———. Proceedings of the twenty-fourth meeting of the Southern Society for Philosophy and Psychology. *Amer. J. Psychol.*, 1929, 41, 505.
46. LARSON, D. L. An experimental critique of the Seashore consonance test. *Psychol. Monog.*, 1928, 38, No. 176, 49-81.
47. LARSON, R. C. Studies on Seashore's "Measures of musical talent." *Univ. Iowa Stud.: Ser. Aims & Prog. Res.*, 1930, 2, No. 6. (First Ser., No. 174.) Pp. 83.
48. LARSON, W. S. Measurement of musical talent for the prediction of success in instrumental music. *Psychol. Monog.*, 1930, 40, 33-73.
49. LENOIR, Z. D. Measurement of racial differences in certain mental and educational abilities. Thesis, Univ. Iowa, 1925.

50. MALMBERG, C. F. The perception of consonance and dissonance. *Psychol. Monog.*, 1918, 25, 93-133.

50*a*. MCCARTHY, D. A study of the Seashore measures of musical talent. *J. Appl. Psychol.*, 1930, 14, 437-455.

51. MCGINNIS, E. M. Seashore's measures of musical ability applied to children of the pre-school age. *Amer. J. Psychol.*, 1928, 40, 620-623.

52. MOOS, J. C. The yardstick applied to musical talent. *Musical Quar.*, 1930, 16, 238-262.

53. MOSHER, R. M. A study of group methods of measurement of sight-singing. *Teach. Coll. Contrib. Educ.*, 1925, No. 194. Pp. 75.

54. MURDOCK, K. A study of differences found between races in intellect and in morality. *School & Soc.*, 1925, 22, 628, 659.

55. MURSELL, J. L. Principles of musical education. New York: Macmillan, 1927. Pp. 300.

56. Music Supervisors National Conference. (*Res. Coun. Bull.*, No. 7.) Chapel Hill: Univ. N. Car., 1927.

57. NORRIS, H. T. A critical review of tests and measurements in music education. *Music Supervisors J.*, 1926, 13, 57-60.

58. OGDEN, R. M. Hearing. New York: Harcourt, Brace, 1924. Pp. xiii+351.

59. PEACOCK, W. A comparative study of musical talent in whites and Negroes and its correlation with intelligence. Thesis, Emory Univ., 1928.

60. PETERSON, J. A functional view of consonance. *Psychol. Rev.*, 1925, 32, 17-33.

61. ———. Comparative psychological study of negroes and whites. *Rep. Ann. Meeting, Div. Anthrop. & Psychol., Nat. Res. Coun.*, 1926.

62. PETERSON, J., & LANIER, L. H. Studies in the comparative abilities of whites and negroes. *Ment. Meas. Monog.*, 1929, No. 5. Pp. vi+156.

63. ROSS, F. B. The measurement of time-sense as an element in the sense of rhythm. *Psychol. Monog.*, 1914, 16, No. 69, 166-172.

64. RUCH, G. M., & STODDARD, G. D. Tests and measurements in high school instruction. Yonkers, N. Y.: World Book Co., 1927. Pp. xvi+373.

65. SALISBURY, F. S., & SMITH, H. B. Prognosis of sight singing ability. *J. Appl. Psychol.*, 1929, 13, 425-439.

66. Schoen, M. The validity of tests of musical talent. *J. Comp. Psychol.*, 1923, 3, 101-121.

67. Seashore, C. E. The psychology of musical talent. Boston: Silver, Burdette, 1919. Pp. xvi+288.

68. ———. Manual of instructions and interpretations for measures of musical talent. New York: Columbia Graphophone, 1919. Pp. 16. (Distributed by C. H. Stoelting Co., Chicago.)

69. ———. A survey of musical talent in the public schools. *Univ. Iowa Stud.: Stud. Child Welfare*, 1920, 1, No. 2. Pp. 36.

70. ———. The inheritance of musical talent. *Musical Quar.*, 1920, 6, 586-598.

71. ———. Measures of musical talent: a reply to Dr. C. P. Heinlein. *Psychol. Rev.*, 1930, 37, 178-183.

72. Seashore, C. E. & Mount, G. H. Correlations of factors in musical talent and training. *Psychol. Monog.*, 1918, 25, No. 108, 47-92.

73. Seashore, R. H. Studies in motor rhythm. *Psychol. Monog.*, 1926, 36, No. 167, 142-189.

74. Smith, F. O. The effect of training in pitch discrimination. *Psychol. Monog.*, 1914, 16, No. 69, 67-103.

75. Spearman, C. The abilities of man. New York: Macmillan, 1927. Pp. xxiii+415.

76. Stanton, H. M. The inheritance of specific musical capacities. *Psychol. Monog.*, 1922, 31, No. 140, 157-204.

77. ———. Psychological tests and musical talent. Rochester, N. Y.: Univ. Rochester, Eastman School of Music, 1925. Pp. 48.

78. ———. Measuring musical talent. *Person. J.*, 1928, 7, 286-292.

79. ———. Seashore measures of musical talent. *Psychol. Monog.*, 1928, 39, No. 178, 135-144.

80. ———. Prognosis of musical achievement. (*Eastman School of Music Stud. Psychol.*, Vol. 1, No. 4.) Rochester, N. Y.: Univ. Rochester, Eastman School of Music, 1929. Pp. 89.

81. ———. Psychological tests—a factor in admission to the Eastman School of Music. *School & Soc.*, 1929, 30, 783-891.

81*a*. Stanton, H. M., & Koerth, W. Musical capacity measures of adults repeated after music education. *Univ. Iowa Stud.:*

Ser. Aims & Prog. Res., 1930, No. 31 (First Series No. 189). Pp. 18.

81*b*. STREEP, R. L. A comparison of white and Negro children in rhythm and consonance. *J. Appl. Psychol.*, 1931, 15, 53-71.

82. TIERNEY, E. M. A study of freshman student achievement in the School of Fine Arts in the University of Nebraska. *Univ. Nebr.*, 1928, 1, No. 2, 54-63.

83. TRAVIS, L. E., & DAVIS, M. G. The relation between faulty speech and lack of certain musical talents. *Psychol. Monog.*, 1926, 36, No. 168, 71-81.

84. VANCE, T. F. Variation in pitch discrimination within the tonal range. *Psychol. Monog.*, 1914, 16, No. 69, 115-149.

85. WEAVER, A. T. Experimental studies in vocal expression. *J. Appl. Psychol.*, 1924, 8, 23-51, 159-186.

86. WILSON, M. E. The prognostic value for music success of several types of tests. *Music Supervisors J.*, 1930, 16, 71-73.

87. WINDHORST, E. L. A study of the sixth-grade attainments stated in the standard course of study adopted by the Music Supervisors National Conference in 1921. Thesis, Univ. Iowa, 1924.

88. WRIGHT, F. A. The correlation between achievement and capacity in music. *J. Educ. Res.*, 1928, 17, 50-56.

UNE ETUDE HISTORIQUE, CRITIQUE, ET EXPÉRIMENTALE DE LA BATTERIE DE TESTS SEASHORE-KWALWASSER

(Résumé)

L'auteur s'est rendu compte d'une façon critique des études antérieures de la batterie de tests Seashore-Kwalwasser. Il a assemblé aussi des données qui sont supplémentaires à certains égards aux études antérieures. On peut résumer ses résultats de la façon suivante:

1. Les "Mesures de Talent Musical" Seashore et les "Tests de Sensitivité Mélodique et Harmonique" Kwalwasser mesurent en général des variables tout différents du comportement.

2. La plus grande partie des membres de la batterie sont bien standardisés. Il semble que les exceptions soient les deux tests Kwalwasser et le test "Sens du Rhythme" Seashore.

3. Chez les élèves, il n'y a presque aucune corrélation avec l'intelligence.

4. Il se montre des différences importantes à l'égard de l'âge dans la plupart des tests.

5. Les différences de sexe sont généralement petites et varient peut-être avec les attitudes envers la musique adoptées par les sujets.

6. On a calculé les quarante-cinq tétrades rendus possibles par les six tests Seashore. Leur médiane est de 0,0286 avec P.E.$_t$ de 0,0212.

7. Les difficultés individuelles trouvées ne sont pas toujours les mêmes attendues en théorie.

8. De grands groupes de sujets ont des difficultés individuelles très semblables.

9. Il se montre de certaines différences de race (surtout nègre-blanche) dans les résultats des tests.

10. Dans toute la batterie seulement le test de "mémoire tonale" et celui de "sens de hauteur" semblent avoir assez de valeur pour un pronostic individuel.

11. On rapporte divers essais de rendre les tests de plus de valeur.

12. Si l'on admet certaines suppositions, la batterie Seashore comme unité aurait une constance de 0,885. Si l'on ajoute les tests Kwalwasser, la valeur devient de 0,870.

13. La plus grande partie des expériences indiquent que la pratique (d'une période d'heures ou de jours à une petite période d'années) n'augmentent pas de beaucoup les résultats Seashore.

14. L'auteur propose "succès dans un conservatoire de musique" comme le meilleur critère dans l'évaluation des tests.

FARNSWORTH

EINE HISTORISCHE, KRITISCHE, UND EXPERIMENTELLE STUDIE ÜBER DIE SEASHORE-KWALWASSER TEST-SERIE

(Referat)

Der Verfasser hat die früheren Studien über die Seashore-Kwalwasser Test-Serie (test battery) kritisch besprochen. Er hat auch Daten zusammen gefasst, welche in gewissen Beziehungen die früheren Untersuchungen ergänzen. Seine Befunde können wie folgt kurz zusammengefasst werden:

1. Die "Masstabe des Musikalischen Talentes" (Measures of Musical *Talent) von Seashore, und die "Masstabe* der Melodischen und Harmonischen Empfindsamkeit" (Tests of Melodic and Harmonic Sensitivity) von Kwalwasser messen im grossen Ganzen weit verschiedene Formen des Benehmens (behavior variables).

2. Die Mehrzahl der Bestandteile der Testserien sind gut standardisiert worden. Es scheinen aber die zwei Kwalwasser Tests und der Test für Rhythmussinn (sense-of-rhythm test) von Seashore hier Ausnahmen zu sein.

3. In Schulbevölkerungen findet sich fast gar keine Korrelation mit "Intelligenz."

4. In der Mehrzahl der Tests zeigen sich wichtige Altersunterschiede (age differences).

5. Die Geschlechtsunterschiede sind gewöhnlich klein, und wechseln vielleicht mit den Haltungen (attitudes) welche die Vpp. der Musik gegenüber annehmen.

6. Die durch die sechs Seashore Tests *ermöglichten* 45 Vierzahlen (tetrads) sind berechnet worden. Ihre Mittelzahl (average) ist .0286, und der wahrscheinliche Fehler .0212.

7. Die festgestellten (relativen) Schwierigkeiten der einzelnen Bestandteilen stehen nicht immer mit den theoretisch zu erwartenden in Einklang.

8. Grosse Gruppen der Vpp. erweisen ziemlich ähnliche Schwierigkeiten mit den einzelnen Bestandteilen.

9. Es zeigen sich gewisse Rassenunterschiede (besonders zwischen Schwarzen und Weissen) in den Testresultaten (test scores). Die Deutung der Resultate ist aber schwierig.

10. Von der ganzen Serie scheint nur der Test des Tongedächtnisses (tonal memory) und der Test der Tonhöheunterscheidung (sense of pitch) zuverlässig genug zu sein, um individuelle Prognose zu ermöglichen.

11. Es wird über mehrere Versuche, die Zuverlässigkeit der Tests zu erhöhen, Bericht erstattet.

12. Werden gewisse Annahmen zugegeben, so erweist die Seashore Serie als Ganzes eine Zuverlässigkeit von .885. Fügt man die Kwalwasser Tests hinzu, so wird der Wert .870.

13. Die Mehrzahl der Experimente deuten an, dass Übung (sich erstreckend über Stunden oder Tagen, oder gar bis über einige Jahre) die Zahlbefunde (scores) bei der Seashore Serie nicht viel ändert.

14. Der Verfasser verteidigt "Erfolg in einem Musikkonservatorium" als das beste Kriterium der Triftigkeit der Tests.

FARNSWORTH

$7.00 per volume
Single numbers $2.00

MONTHLY
Two volumes per year

June, 1931
Volume IX, No. 6

GENETIC PSYCHOLOGY MONOGRAPHS

Child Behavior, Animal Behavior, and Comparative Psychology

A COMPARISON OF DIFFICULTY AND IMPROVEMENT IN THE LEARNING OF BRIGHT AND DULL CHILDREN IN REPRODUCING A DESCRIPTIVE SELECTION*

From the State Teachers College, Buffalo, New York

By

F. T. WILSON

*Accepted for publication by Leta S. Hollingworth of the Editorial Board and received in the Editorial Office, August 20, 1930.

Worcester, Massachusetts

Entered as second-class matter December 1, 1925, at the postoffice at Worcester, Mass., under Act of March 3, 1879

CONTENTS

I

INTRODUCTION

In a previously published study (1) the learning of bright and dull children was compared in the task of reproducing a descriptive selection. It appeared from that study that for two groups of children, bright and dull, respectively, whose average mental ages were about ten years and six months each, the results were practically the same, although the average chronological ages of the groups were about nine and twelve years, each. A third group composed of bright twelve-year-old subjects scored much higher than these two groups, and a fourth group of dull nine-year-old children scored much lower. These results were to have been expected, particularly as the selection of the subjects was made on the basis of IQ scores as measured by the Stanford Revision of the Binet-Simon Test—a test whose correlation with language abilities is said to be very high.

This report gives the result of further study of the four groups in this task, in regard to the ranking of certain "idea" units and to improvement during a series of practices in their reproduction. The data for the subject groups are given in Table 1. Three more cases are included in these data than were used in the previous study.

The selection used for the task was the following adaptation and extension of Test 11 of the Herring Revision of the Binet-Simon Tests. The lines cross-

TABLE 1
DESCRIPTION OF SUBJECTS

	Dull 9	Bright 9	Dull 12	Bright 12
Number of cases	15	17	16	15
Average CA	9-1	9-2	12-2	12-2
CA range	8-11 to	9-4	12-0 to	12-5
Average MA	7-9	10-5	10-4	13-11
Average IQ	85	114	85	115

ing the print indicate the division into "ideas" as counted in the tests—a total of 48.

"In a little / cottage / on the banks / of a river / in France / there lived a farmer / and his wife. / Their home was / in a pleasant spot; / behind it / was a forest; / in front / was a maple / tree. / Under its branches / there sat their daughter / a good / and gentle girl, / whose work it was / to look after the sheep / of her father's / flock. / The name of Goldie / might have been given her / as / her hair was golden / brown, / while / her red / cheeks, / smiling / dark / eyes / and bright / lips / gave one delight / to behold. / But / she was called Gloria / by her father, / Sunshine / by those who knew her / well, / and Fairy / Girl / by her mother's / little / sister. /"

The experiment covered five successive days, on each of which the selection was read to the child by the examiner and then immediately given back by the subject. On each of the fourth and fifth days a retention test was made before the selection was read to the subjects, after which the usual reading and reproduction followed. The procedure of the experiment is outlined in the study above cited from which the following extract is taken.

"The subject was directed to listen carefully to a 'story' so that he could retell it. The experimenter read the selection aloud, distinctly and with little emphasis or

expressive interpretation. At the conclusion the subject was directed to tell everything he could remember. There was no time limit, each subject being allowed as much time as he apparently could use to any purpose. In case very little was reproduced encouragement was given by such direction as 'Go ahead,' or 'What else?' In a few cases where the results were particularly meager, urging was continued, but nearly always with not a single added idea. Apparently after a minute or so nothing further came to mind, either for the fast or for most of the slow reproducers.

"The record was the number of ideas correctly given, identical wording not being required. When words of the selection were incorrectly used, however, they were not counted, as, for instance, 'her father called her Goldie' (not correct). 'Goldie', the word incorrectly used was not counted. On the other hand, when single words, such as 'river,' 'cottage,' and the like, were given—as happened with a few of the younger dull subjects—they were counted, if they were important words in any idea group. This concession affected the scores of but the dull nine-year-olds and can only slightly have reduced the differences between them and the other groups, differences which, with the procedure followed, are very striking" (*1*, pp. *15-16*).

Table 2 gives the total average number of ideas reproduced for all subjects during all trials of the experiment.

The purpose of the present study is to analyze the data summarized in Table 2, in an attempt to find comparisons of difficulty among the four subject

TABLE 2

Total Average Number of Ideas Reproduced

Dull 9	Bright 9	Dull 12	Bright 12
99.7	180.4	176.2	235.6

TABLE 3
BASIC TABLE OF DERIVED SCORES FOR EACH GROUP OF SUBJECTS FOR EACH IDEA

	Idea	Scores Dull 9	Bright 9	Dull 12	Bright 12	Total
1	*In a little*	6	6	6	7	25
2	cottage	4	6	5	7	22
3	on the banks	4	4	6	9	23
4	of a river	4	5	7	9	25
5	in France	5	8	8	9	30
6	lived a farmer	4	5	5	4	18
7	and his wife	4	5	6	7	22
8	Their home was	2	4	4	4	14
9	in a pleasant spot	2	5	4	6	17
10	behind it was	4	8	8	8	28
11	*a forest*	4	8	8	8	28
12	in front was	3	8	8	8	27
13	a tree	6	8	8	9	31
14	maple	6	8	7	9	30
15	*Under its branches*	5	8	6	8	27
16	sat their daughter	6	8	8	8	30
17	a good . . girl	2	3	4	3	12
18	and gentle	1	2	1	3	7
19	whose work it was	4	5	4	7	20
20	to look after the sheep	6	7	7	8	28
21	of her father's	2	4	4	6	16
22	flock	3	5	5	7	20
23	The name of Goldie	2	5	6	9	22
24	might have been given her	0	5	6	8	19
25	as	0	4	6	6	16
26	her hair was golden	4	5	7	8	24
27	brown	2	5	5	5	17
28	while	0	0	2	1	3
29	her red	1	2	2	6	11
30	cheeks	2	2	3	7	14
31	smiling	0	1	2	4	7
32	dark	0	1	2	5	8
33	eyes	0	2	3	6	11
34	and bright	0	1	2	4	7
35	lips	0	1	2	4	7
36	gave one delight	0	1	1	2	4
37	to behold	0	1	1	2	4
38	But	0	2	2	4	8
39	she was called Gloria	4	7	6	9	26
40	by her father	4	7	7	9	27
41	Sunshine	3	7	6	8	24
42	by those who knew her	3	8	6	9	26
43	well	2	6	3	6	17
44	and Fairy	2	6	4	7	19
45	Girl	2	5	5	6	18
46	by her mother	2	7	5	8	22
47	sister	2	6	5	8	21
48	little	2	5	4	7	18

groups. This will be done in two ways:

1. By study of difficulty as indicated by rank of ideas according to scores derived from the number of correct reproductions.

2. By study of gains as indicated by differences between first and last scores.

The derivation of the scores used in the tables of this study is shown in the following formula:

$$\frac{(\text{Number of times ideas were reproduced} \times 100)}{(\text{Number of trials} \times \text{number of subjects})} \div 10$$

Fractional remainders were dropped. This score avoids minute distinctions—not warranted by the nature of the experiment— and reduces figures to those on a scale of 10.

II.

DIFFICULTY

Table IV*a* gives the order of ideas, when ranked according to the total scores made by all four groups combined.

A difference of one in the scores stands for a difference of about ten per cent, so the rank position given in this table probably represents differences of some significance. The range, with a possible score of 40, is from 31 down to 3, descending with quite uniform regularity. The position of the idea in the selection is given by the numbers in the right-hand column of the table.

Study of the data soon suggested that a grouping of the 48 ideas into larger groups would be desirable. Accordingly, 16 larger-idea groups were determined. Their selection was made on two grounds: (*a*) the scores for each idea, and (*b*) logical relations of ideas. It may appear that such combinations are not justified. It seemed out of the question, however, to depend upon but one of these factors. The score results gave evidence beyond question that certain ideas were reproduced well and others very poorly. In certain instances the scores showed that, although ideas logically belonged together, it could not be said that they had been equally well learned by the subjects. On the other hand, certain ideas, having nearly the same scores, were so widely separated in position in the selection that it did not seem correct to consider that they belonged together in a larger idea unit. The

TABLE IVa
Ranking of Ideas by the Grand Total of All Four-Group Total Scores

Rank	Score	Idea	Position number of idea in the selection
1	31	tree	14
2	30	in France	5
3	30	was a maple	13
4	30	sat their daughter	16
5	28	behind it	10
6	28	was a forest	11
7	28	to look after the sheep	20
8	27	in front	12
9	27	under its branches	15
10	27	by her father	40
11	26	she was called Gloria	39
12	26	by those who knew her	42
13	25	cottage	1
14	25	of a river	4
15	24	her hair was golden	26
16	24	Sunshine	41
17	23	on the banks	3
18	22	In a little	2
19	22	and his wife	7
20	22	the name of Goldie	23
21	22	by her mother's	46
22	21	sister	48
23	20	whose work it was	19
24	20	of her father's	21
25	19	might have been given her	24
26	19	and Fairy	44
27	18	there lived a farmer	6
28	18	Girl	45
29	18	little	47
30	17	in a pleasant spot	9
31	17	brown	27
32	17	well	43
33	16	flock	22
34	16	as	25
35	14	Their home was	8
36	14	cheeks	30
37	12	a good . . girl	17
38	11	her red	29
39	11	eyes	33
40	8	dark	32
41	8	But	38
42	7	gentle	18
43	7	smiling	31
44	7	bright	34
45	7	lips	35
46	4	gave one delight	36
47	4	to behold	37
48	3	while	28

following table gives these larger idea groups in their score order. Their positions in the selection are indicated and their combined average scores stated.

Three of the original 48 ideas—brown, But, and while—were not combined with any others. There seemed to be no other logically related ideas of comparable scores to which they could have been attached. To have placed them in groups where they were logically related would have covered up score differences which it seems desirable to have appear.

The rankings shown in these tables make comment permissible. That the differences in scores represent only differences in the nature of the difficulty of the

TABLE IV*b*

RANKING OF LARGER-IDEA GROUPS BY THE AVERAGE OF THE GRAND TOTAL OF FOUR-GROUP TOTAL SCORES

Rank position	Small-idea selection number	Score	Idea group
1	13–16	29.6	a maple tree, Under its branches sat their daughter
2	10–12	27.7	Behind it was a forest. In front was
3	3–5	26.0	on the banks of a river in France
4	1–2	23.5	In a little cottage
5	39–48	21.9	she was called Gloria by her father, Sunshine by those who knew her well and Fairy Girl by her mother's little sister.
6	19–22	21.1	whose work it was to look after the sheep of her father's flock
7	23–26	20.4	The name of Goldie might have been given her, as her hair was golden
8	6–7	20.0	lived a farmer and his wife
9	27	17.0	brown
10	8–9	15.5	Their home was in a pleasant spot
11	29–30	12.5	red cheeks
12	17–18	9.5	a good and gentle girl
13	31–35	8.0	smiling dark eyes and bright lips
14	38	8.0	But
15	36–37	4.0	gave one delight to behold
16	28	3.0	while

ideas themselves is probably not the whole explanation of the score differences. Yet, in certain cases, at least, there must have been differences in the nature of the difficulty of the ideas themselves.

Obvious and outstanding is the fact that there are large score differences. Comparing the scores for either grouped or small ideas, it is evident that those near the bottom are much lower than those near the top. Among the ideas at the bottom are "But" and "while." Perhaps these ideas are really harder for children than the others. The words are abstractions indicating certain relationships between other ideas and are known to appear comparatively late in the vocabularies of young children.

"Gave one delight to behold," rank 15 in the list of the 16 larger-idea groups, is rather archaic or stilted and probably is quite unusual in the language expression of children. It also is abstract to a considerable degree. The relation between being presented with something—especially such a thing as "delight" —when all one does is to behold, might well stump the directness of a child's mind. It seems reasonable to hold that this idea had inherent difficulty.

"Their home was in a pleasant spot," rank 10, also sounds strange. A home has rather vague but real extension in size. A spot, relatively, is tiny. Also, it seems conceivable that, with a life of more or less continual learning about spots being nothing if not unpleasant, there might have been some blocking when it came to saying that "Their home was in a pleasant spot!"

Certain qualities inherent in the ideas themselves seem to *merit consideration as factors in difficulty.* Such ideas are in themselves relatively so difficult that they are not sufficiently well grasped to be given back in reproduction.

There are other groups of ideas, however, which also rank low and should be considered in order to find some explanation for their low scores. "A good and gentle girl" stands in rank position 12, in the larger group list of 16. The question suggests itself as to whether this conception is difficult to comprehend in the minds of children, or if it is uninteresting, unimportant, or perhaps unpopular among them! The idea "gentle" may be unusual, but certainly not "a good girl." The position of this phrase is immediately following the words, "sat their daughter," part of the larger-idea group in sixth place. It does not seem correct to account for the low position of the idea in terms of its *inherent difficulty. Other factors than* difficulty of the ideas must have been operating.

The remaining ideas which are low in the rank-order list are "red cheeks" and "smiling dark eyes and bright lips." Two conditions may explain, at least in part, their low scores. The words themselves are not unusual to children, but the use of the adjectives with the particular nouns they modify in these cases may be said to be unusual. Usually lips are smiling and red and eyes are bright. The unusual combination was made with intent. It was thought that possibly such unaccustomed use might result in distinguishing bright and dull. Later comment upon that expected differ-

ence will be made. The other condition which may explain the low rank order of these ideas is that of their numerical difficulty. Six different adjectives of personal description are used in this part of the selection, all in one sentence. Perhaps memory-span ability explains the matter. Of course, reproduction of the whole selection is related to memory span, but it seems fair to consider that that particular function of reproduction might explain the low score, where so many ideas of the same kind are in sequence. The idea "brown," which is only at about the middle point of the score distribution, may owe its rank in part to this same factor, as it is part of the sequential description.

Turning to the smaller-idea groups which rank high, it is found that "in France" stands along with "sat their daughter" and "was a maple" "tree" at the top of the list. Perhaps familiarity, due to much experience, explains the high positions of these ideas. Such an explanation may be acceptable for the latter three. It does not clearly account for the position of "in France." It is doubtful if any of the children had ever seen France. The great extent to which France occupied the American mind because of the World War may have had some effect, although the oldest subjects of the experimental groups were but four years old when the War ended. "Was a forest" also stands at the top of the list. The children of the experiment probably never saw a forest. Part of them lived in a Jewish orphanage in New York City and the others in a community of Italian, Polish, and American people in Buffalo, New York. Familiarity of

the ideas was probably one factor, but not the chief one, affecting the scores of these ideas. What these may have been will be discussed briefly.

It may fairly be said of these other factors that they have to do with subtle relationships which the ideas of the selection made as it was heard and fitted into the mental life of the children. Chief among them must have been those which picked out the main idea of interest to the children around which they built up a group of ideas into a synthetic whole. It is significant that the idea around which the reproductions were built was "daughter" and not "girl," a single idea which stands very low. It was not a "good and gentle girl," but a "daughter," which held the center of the stage.

A second factor which seems important in explaining the order of rankings of the ideas is the high position of those ideas which relate to location. "Where" seems to precede all concepts except "what." Perhaps to children, as to rats and guinea-pigs, ideas of location come early in the integration of experience and are fixed deeply.

Further analysis of causes and effects will not be hazarded at this point. Briefly, the rank order of ideas shows that the heroine in her setting emerges as the most important, or the most interesting, or the easiest, of the larger-idea groups. More about her setting follows. Then come her name, her work, and lastly, her attributes and their audience effect. Neither beauty of cheek, eyes, nor lips, nor virtue of the character of the heroine seem to have mattered as much to

the 60 odd boys and girls of the experimental groups as did her location under a maple tree, beside a forest over in France.

The next question is: Are there differences in rank order for the 48 ideas or for the 16 larger-idea groups among the four subject groups?

These facts in regard to the 48 ideas are shown by means of quartile positions. The scores for each group, as given in the basic table, were put into frequency tables and the quartile positions for each found. Table 5 gives this result. Part *"a"* gives the 16 ideas which were found to be in the same quartile in all four subject groups. Part *"b"* gives 22 ideas found to be in the same or the next lower quartile. Part *"c"* gives the remaining 10 ideas which varied two or more quartile positions in the four groups.

TABLE 5

QUARTILE POSITIONS OF THE FORTY-EIGHT IDEAS FOR ALL GROUPS

Number of idea	Idea	Quartile positions Dull 9	Bright 9	Dull 12	Bright 12
	Part a, same quartiles				
5	in France	4	4	4	4
11	a forest	4	4	4	4
13	maple	4	4	4	4
14	tree	4	4	4	4
16	sat their daughter	4	4	4	4
41	Sunshine	3	3	3	3
21	father's	2	2	2	2
18	a good . . . girl	1	1	1	1
28	while	1	1	1	1
31	smiling	1	1	1	1
32	dark	1	1	1	1
34	and bright	1	1	1	1
35	lips	1	1	1	1
36	gave one delight	1	1	1	1
37	to behold	1	1	1	1
38	But	1	1	1	1
Total					16

TABLE 5 (*continued*)

QUARTILE POSITIONS OF THE FORTY-EIGHT IDEAS FOR ALL GROUPS

Number of Idea	Idea	Quartile positions Dull 9	Bright 9	Dull 12	Bright 12
	Part b, same or next quartile positions				
4	of a river	4	3	4	4
10	behind it was	3	4	4	4
12	in front was	3	4	4	4
15	under its branches	4	4	3	4
20	to look after the sheep	4	4	4	3
39	she was called Gloria	4	4	3	4
40	by her father	4	3	4	4
42	by those who knew her	3	4	3	4
7	and his wife	3	3	3	2
9	in a pleasant spot	2	3	2	2
19	whose work it was	3	2	2	3
22	flock	3	3	2	3
44	and Fairy	2	3	2	2
45	Girl	2	3	2	2
47	little	2	3	2	3
48	sister	2	3	2	2
8	Their home was	2	2	2	1
17	good . . . girl	2	2	2	1
27	brown	2	2	2	1
29	red	1	1	1	2
30	cheeks	2	1	2	2
33	eyes	1	1	2	2
Total					22
	Part c, ideas which vary two or more quartile positions				
(1)	Low dull 9 and high bright 12				
23	The name of Goldie	2	2	3	4
24	might have been given her	1	2	3	4
25	as	1	2	3	2
46	by her mother's	2	3	2	4
(2)	High dull 9 and low bright 12				
1	In a little	4	3	3	2
2	cottage	4	3	2	2
6	lived a farmer	4	2	2	1
(3)	Unclassified				
3	on the banks	4	2	3	4
43	well	2	3	1	2
26	her hair was golden	3	2	4	3
Total					10

The outstanding fact presented by these tables seems to be that, in each of the four subject-group distributions, 38 of the 48 ideas are found to be in the same

quartiles or in adjacent quartiles. Apparently for the four groups of children there is not much difference, if any, in relative difficulty of these 38 ideas.

Of the remaining 10, those placed in Part *c* (3), and described as "unclassified," suggest no other explanation for their varying quartile positions than that of chance. In the case of idea 3, "on the banks," bright-nine group places the idea in next to the lowest quartile, the other in quartiles 3 or 4. In the case of idea 43, "well," dull-twelve drops the idea down into the lowest quartile, while the others place it in quartiles 2 or 3—or one may say that bright-nine raises the idea to quartile 3. In idea 26, "her hair was golden brown," dull-twelve raises it to the highest quartile—or one may say bright-nine drops it to next to the lowest. There seem to be no logical reasons for these variations in position except that of chance.

This leaves seven ideas to examine. Ideas 23, 24, and 25, "The name of Goldie might have been given her, as," seem to have a logical as well as an adjoining sequence. This clause and logical connective word belong to that part of grammar which ordinarily comes near the back of the book. The thought is contingent, but expresses a possibility which, as a matter of fact, is contrary to the facts of the case. It is, therefore, an abstraction in a much more involved way than is the plain straight-forward account of things which are facts. Such an explanation may account, in part, for the differences scored by dull-nine and bright-twelve. It also suggests a reason for the nearly same position of the ideas for bright-nine and dull-twelve who have the

same mental age. Ability to reproduce abstract ideas dependent upon other given factors is a measure, perhaps, of mental development.

The remaining idea of this set—"by her mother's"—does not have as clear a case of progressive rise in quartile position from dull-nine to bright-twelve, but seems to merit some consideration in this regard. Reference to the selection shows that this phrase is one of three descriptive ideas relating to the name "Fairy Girl," which, in its turn, is one of three ideas resting upon the idea "she was called." Just why this particular one of these related nine ideas—40 to 48—should be the one ranked high in position by the bright-twelve and low by the dull-nine is not clear. The tables show that number 40 and number 42 are in quartiles 3 and 4 for all groups, 41 in quartile 3 for all groups, and 44, 45, 47, and 48 in quartiles 2 or 3. Number 43 has been disposed of by falling back upon chance. Perhaps chance would be the best recourse for this case, also. If the quartile position for the bright-twelve had been three instead of four, no attempt would have been made to explain a supposed difference.

There remain now only three ideas whose variation in quartile position among the four groups requires discussion. The dull-nine found these three ideas—"in a little," "cottage," "lived a farmer"—so manageable that they fall in the highest quartile for that group. Were they so hard for the bright-twelve that three-fourths of the 48 ideas out-ranked them? Hardly possible. The differences are too apparent to permit calling upon chance. Were they too simple, easy, uninterest-

ing? Did other ideas catch the imagination more forcefully in the minds of these brighter, older children? This seems more likely to be the truth of the matter. Perhaps, too, these ideas belong less genuinely to the selection than those more directly connected with the heroine. *Certainly it cannot be maintained* that there is an inherent difficulty in the nature of these ideas making them more difficult for the older and brighter children.

CONCLUSION 1

From this study of the quartile positions of the 48 ideas of the selection, it seems that one set of three—"The name of Goldie might have been given her, as"—had such a degree of abstractness that it was relatively much harder for young and dull, than for the bright twelve-year-old subjects. *Possibly* this very characteristic stimulated the attention of the brighter group.

CONCLUSION 2

From this study of quartile positions of the 48 ideas of the selection, it seems that factors other than the in*herent difficulty of the ideas themselves must* have been at work in the reproduction given by the children. One such factor may have been intellectual interest, as suggested above. The contrary may, however, have been the factor—boredom and ennui. Possibly a subtle factor of a sense of coherence was also operating in the reactions of the bright twelve-year-old subjects.

How do the ideas rank when grouped into the 16 larger-idea groups? This question may be studied

TABLE 6
QUARTILE POSITIONS OF LARGER-IDEA GROUPS

Ideas	Dull 9	Bright 9	Dull 12	Bright 12
13-16	4	4	4	4
1- 2	4	4	3	3
3- 5	4	3	4	4
6- 7	4	3	3	2 ?
19-22	3	3	3	3
10-12	3	4	4	4
39-48	3	4	4	4
8-9	3	2	2	2
27	2	3	2	2
17-18	2	2	2	1
23-26	2	2	4	4 ?
29-30	2	2	2	3
28	1	1	1	1
31-35	1	1	1	2
36-37	1	1	1	1
38	1	1	1	1

from the data in Table 6. The question marks indicate the variations of interest in our inquiry. These variations present nothing new. The discussion given in regard to the smaller-idea groups seems to cover the matter with the larger-idea groups. Attention is called to the apparent rise in position of the idea group 36-37 —"gave one delight to behold"—which takes the place of ideas 17-18—"a good and gentle girl."

Comment is in order at this point in regard to the result anticipated in connection with the six somewhat juggled adjectives. There is no evidence that the bright were, relatively to the other ideas, more capable of handling these ideas than were the dull. For all subject-groups these ideas were difficult, falling either in the lowest or in the next to the lowest quartiles. The same is true, moreover, of the other ideas ranked low

in the table of total-score rankings, namely: "gave one delight to behold," "Their home was in a pleasant spot," and "a good and gentle girl." They seem to be just as hard, relatively to the other ideas, for the bright as for the dull and for the older as well as for the younger.

CONCLUSION 3

Dull and younger subjects handled a long series of descriptive terms, in which were several "juggled" ad-

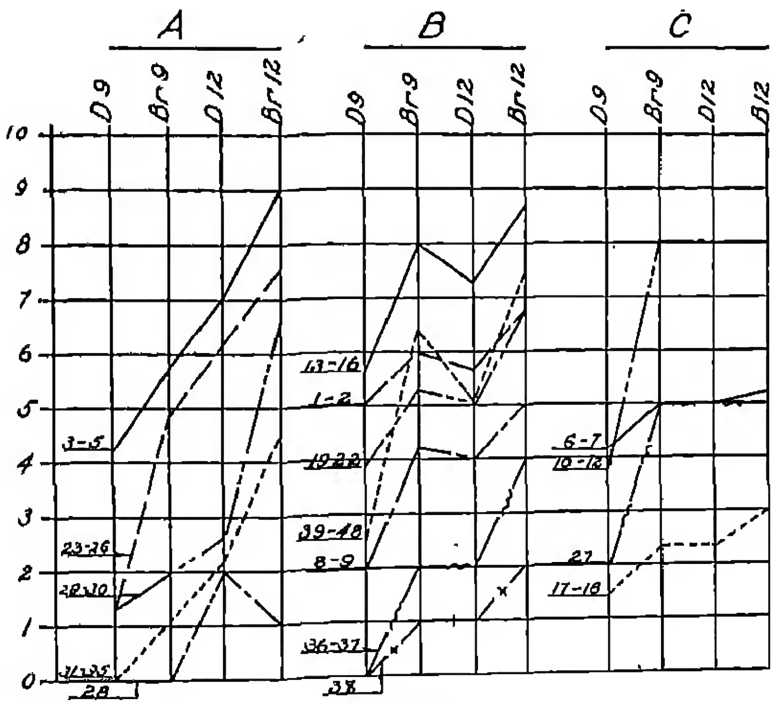


FIGURE 1

RELATION OF SUBJECT-GROUPS ACCORDING TO TOTAL SCORES MADE ON LARGER-IDEA GROUPS

jectives, as well, relative to the other ideas of the selection, as did the bright and older subjects.

A relationship among the four subject groups in regard to idea scores is shown by the graphs in Figure 1, drawn from the data of Table 3. In this figure there are pictured the total scores made by each subject group for the 16 larger-idea groups. Graph A shows those idea-groups which tend to have higher scores for each subject group as the groups are arranged in order from dull-nine to bright-twelve. The graph suggests that, for these ideas, age had about as much to do with higher scores as had brightness. The five ideas belonging together because of this relationship are:

3- 5	on the banks of a river in France
23-26	The name of Goldie might have been given her as her hair was golden
28	while
29-30	her red cheeks
31-35	smiling dark eyes and bright lips

Idea number 28, "while," is included in this set with the belief that its true score for the bright-twelve is nearer 3 or 4, and that the found score is due to chance.

In the second set of ideas shown by Graph B dull-twelve children score slightly less than, or more probably about the same as, the dull-twelve. This relationship suggests that, for these ideas, brightness is a more effective factor than is age. These ideas are:

1- 2	In a little cottage
8- 9	Their home was in a pleasant spot
13-16	was a maple tree. Under its branches sat their daughter
19-22	whose work it was to look after the sheep of her father's flock

36-37 gave one delight to behold
38 But
39-48 she was called Gloria by her father, Sunshine by those who knew her well and Fairy Girl by her mother's little sister.

In the third set of ideas, shown by Graph C, bright nine-, dull twelve-, and bright twelve-year-old children scored practically the same. This suggests that other factors than age or brightness were operating. These ideas are:

6- 7 lived a farmer and his wife
10-12 behind it was a forest, in front
17-18 a good and gentle girl
27 brown

This classification of large-idea groups, according to the differences in the curves of score results by the four subject groups, results in a rough division into two groups; first, the more difficult ideas in the group shown by Graph A; second, the easier ones in the group shown by Graph B. Perhaps the two ideas shown at the bottom of Graph B really belong in Graph A. The like scores made by bright-nine and dull-twelve may be distortions of the real facts. The same may be true in the case of ideas 17-18, shown in Graph C. Without doing violence to the criteria by which the graph reproductions were made, all of the larger-group ideas, scores for which were less than half that of the highest score made, will be classed in Graph A. Only one other idea-group—3-5—will be found among them. This group stands high in position and score. Why it should appear with the others as classified in Figure 1 seems inexplicable from the data available.

Shifting the three idea-groups above indicated—38, 36-37, and 17-18—to Graph A leaves Graph B with five of the large-idea groups, all of which had scores above 15 on the score range of 3.0 to 29.6.

There are left, then, three idea-groups in Graph C. Can the two groups—10-12 and 6-7—be shifted to Graph B where the other easier-idea groups are found? The first of these two ideas might well be said to belong there and to owe its same score by dull-twelve and bright-twelve to the fact that the scores made by these two subject-groups were so near the top of the range that little or no difference would be expected. In regard to the second idea-group, 6-7, the factor previously suggested—the unimportance of the idea, "the farmer and his wife," in relation to the central theme of the selection—may well explain why the bright twelve-year-old children neglected this part of the reproduction. This leaves the idea 27—"brown"—to explain. The only suggestion to be made here is that chance factors operated to lower the true score of the bright-twelve.

With or without accepting the above-suggested shifts of certain of the idea-groups, shifts which seem entirely warranted in consideration of the complexity of the data, classification of the 16 larger-idea groups into two divisions, explainable in terms of difficulty as measured by scores, seems to justify two further conclusions:

Conclusion 4

In regard to the more difficult ideas as indicated by scores, age as well as brightness is an effective factor.

CONCLUSION 5

In regard to the easier ideas as indicated by scores, brightness is a more effective factor than age.

Other factors in explanation of the results will be briefly discussed. The position of ideas in the selection may be a factor in their scores in reproduction. Unfortunately, the experiment provided no direct check upon this factor. It is possible, however, to investigate the question empirically by noting the relation of position and scores. Table 4*a* shows that the highest scores were not made by the ideas first in position, nor the lowest scores by those at the last of the selection. There is, however, an apparent tendency for the ideas in the early part of the selection to score toward the top of the list. Further analysis casts doubt upon the influence of position as a factor in the scores. If the total scores are reduced to percentages for each of the four subject-groups, as shown in Table 7, it is found that the first half of the selection was much easier for the dull-nine than was the last half. The same is true, but not so much so, for bright-nine and dull-twelve. The percentage for bright-twelve might mean no real difference, certainly at least a very slight difference between the two halves.

Previous discussion has already suggested that

TABLE 7

PERCENTAGE OF TOTAL SCORES MADE IN EACH HALF OF THE SELECTION

Dull 9	Bright 9	Dull 12	Bright 12
72	65	65	55

factors other than position in the selection operated to keep down scores for certain ideas. The ideas demanding abstract ability and a long memory span are in the last half of the selection. Those selecting the heroine and locating her in space come in the first half. The large last group of ideas, giving the names of the daughter, ranked high. It seems improbable that "final spurt," "last impression," or anything related to such position in the series, explains this high score nearly as well as do the partially known factors of subtle relationships already proposed.

Still another factor suggested to explain difference in score is that of the part of speech concerned. There seems to be little of importance in this regard with the selection used. The conjunctions, "But", "while", and "as" rank low, but, aside from them, there seems to be no relation between score and part of speech. As far as nouns, adjectives, and verbs are concerned, they are spread from the top to the bottom of the rank-order list.

Conclusion 6

Position of ideas in the selection and part of speech seemed to have little if any influence upon scores.

III

IMPROVEMENT

Table 8 gives the data from which improvement has been studied. The scores are the averages of the basic scores for each idea for Trials 1 and 2, called the first scores, and the averages of the basic scores for each idea for the last retention trial and the last reproduction trial, called the last scores. This table has been

	Low	Middle	High	Total
Br > D	18 34 29 35 30 36 31 37 32 33 $\boxed{10}$	1 8 43 47 $\boxed{4}$		14
Same	17 21 28 38 $\boxed{4}$	2 23 41 4 24 42 6 25 44 7 27 45 9 39 46 22 40 48 $\boxed{18}$	5 13 14 15 16 20 $\boxed{6}$	28
D > Br		3 19 26 $\boxed{3}$	10 11 12 $\boxed{3}$	6
Total	14	25	9	48

FIGURE 2

Classification of Ideas According to Relation of Bright and Dull in Gains Made and to Approximate Positions of Scores in the Range

used as source for the data which Figure 2 presents. In this diagram the dull are compared with the bright, dull-nine and dull-twelve having been combined, and likewise bright-nine and bright-twelve. The classification of the ideas in the diagram is roughly distinct. Each idea is placed according to two factors: first, its approximate position in the range of scores—in the column headed "Low" if the score is low, in the column headed "Middle" if the score is around that part of the range, and in the one headed "High" if the score is high; secondly, each idea is placed in one of the three rows, according to the relation of scores of dull and bright groups: in the first row, if the bright made more improvement than the dull; in the second row, if the improvement was about the same; in the third row, if the dull groups made more improvement than the bright.

The diagram shows that, with more than half of the ideas, the dull and bright groups gained about the same. Of the 28 ideas so related, 4 had low scores, 6 had high scores, and 18 had scores near the middle of the range. That is, for these ideas there seems to be no clear relation of amount of gain to size of score.

A second group to examine is that of the six ideas for which the gain was more by the dull than by the bright. Three of these ideas had high scores, and three had scores near the middle. The three in the high column may be accounted for by the fact that the bright group scored so high on them in the first trials that comparatively little gain was possible, while the dull, although scoring high on these ideas, began with

TABLE 8

Basic Table of First and Last Derived Scores

	Idea	Dull 9		Bright 9		Dull 12		Bright 12	
		F	L	F	L	F	L	F	L
1	In a little	6	6	5	7	7	5	7	7
2	cottage	4	5	5	7	4	5	6	7
3	on the banks	2	5	3	5	4	8	7	10
4	of a river	3	6	4	7	6	8	8	10
5	*in France*	4	6	5	8	6	9	8	10
6	lived a farmer	3	4	3	6	4	7	3	5
7	and his wife	2	4	5	6	4	6	4	8
8	Their home was	1	3	2	6	2	6	1	8
9	in a pleasant spot	0	4	5	6	2	6	3	8
10	behind it was	3	5	8	9	6	9	8	9
11	a forest	3	6	8	9	6	9	7	9
12	in front was	1	5	8	9	6	9	7	9
13	a tree	4	6	8	10	6	9	7	9
14	maple	5	6	7	9	7	8	7	10
15	Under its branches	4	6	5	9	5	7	8	9
16	sat their daughter	6	6	7	8	8	8	8	9
17	a good . . girl	1	3	1	5	2	5	2	5
18	*and gentle*	0	2	0	4	0	3	2	6
19	whose work it was	2	6	3	5	3	6	7	9
20	to look after the sheep	5	7	6	8	5	8	6	9
21	of her father's	2	3	4	5	3	5	3	7
22	flock	2	4	4	5	4	6	5	8
23	The name of Goldie	0	2	5	6	5	6	8	9
24	might have been given her	0	1	3	7	4	7	8	9
25	as	0	2	3	6	4	6	5	7
26	her hair was golden	3	5	4	5	6	8	8	9
27	brown	1	3	3	6	4	6	4	7
28	*white*	0	0	0	0	1	3	1	1
29	her red	1	1	1	3	1	3	5	8
30	cheeks	1	2	1	3	3	4	6	8
31	*smiling*	0	0	0	3	2	3	3	6
32	dark	0	1	1	2	1	3	4	7
33	*eyes*	0	1	1	5	2	4	5	8
34	and bright	0	1	0	2	1	2	2	5
35	*lips*	0	1	0	3	2	3	2	5
36	gave one delight	0	0	0	2	0	1	0	6
37	to behold	0	0	0	2	0	1	1	5
38	But	0	0	2	3	1	2	4	5
39	she was called Gloria	4	4	6	8	5	6	9	10
40	by her father	4	6	6	7	6	6	9	10
41	Sunshine	2	3	5	8	5	7	8	9
42	by those who knew her	1	3	6	9	3	6	8	9
43	well	1	2	3	7	2	4	3	7
44	and Fairy	0	3	4	8	2	6	5	9
45	Girl	1	2	3	7	2	7	3	8
46	by her mother's	0	3	4	9	2	7	7	10
47	sister	1	4	3	8	2	6	5	10
48	little	0	4	3	9	1	7	4	9

considerably lower scores than did the bright, and accordingly were able to and did gain more than the others. The same explanation seems to account for idea 26, placed in the middle of the range column. Bright-twelve scored high on that idea in the first trials and so could gain but little, and yet the scores of the other three groups make it seem necessary to place the idea in the middle column instead of in the high. Both dull groups were able to gain 20% on the scale, while bright-nine, for some chance reason, made only 10%. The differences are not large, at any rate.

Two ideas remain, 3 and 19. In these, part of the explanation for the greater gain of the dull groups may be, again, the high first scores of the bright-twelve group. But the main arithmetical reason is that bright-nine gained only 20% in each case, although beginning below the middle of the range, while the dull groups made 30% and 40% each on both of these ideas. Given just about the same chance to make improvement, why did the bright-nine group fail to gain roughly as much as the two dull groups on the ideas "on the banks" and "whose work it was"? Perhaps there's a dissertation in that—on chance!

In 14 cases the bright groups gained more than the dull. Ten of these were in the low part of the range. The other four were in the middle of it. Examination of the latter group brings out certain interesting facts. Idea 1, "In a little," had no gains by either dull-nine or bright-twelve. Dull-twelve lost 20%—the only case where any loss in score occurred. All the scores were high-middle scores—first, last, dull, and

bright. The moderate gain by the bright-nine and this peculiar loss by dull-twelve put the idea in the class of those in which the bright gained more than the dull. It probably does not belong there but in the next group, where gains are about the same. The other three may very well be said to show a difference in ability to gain between bright and dull groups. What that may be will be suggested in connection with the examination of the ten ideas in the low score classification given in the next paragraph.

All the ideas in the group of low scores and more-gain-by-bright-than-dull-subjects seem to belong there. In every case the gains are much more for the bright than for the dull. The striking thing about these ideas is that they include all of a series of descriptive material, namely, the series, "her red cheeks, smiling dark eyes and bright lips, gave one delight to behold." For some reason this part of the selection was harder for the dull than for the bright as far as improvement is concerned. But it was also harder than other ideas in total scores for all the groups. It has already been noted that it is these ideas which were in the lowest or in the next to the lowest quartiles for all four groups and at the bottom of the table of rankings by total scores. It would seem that the bright subjects were able to make greater gains than the dull in ideas relatively the hardest, for some reason inherent in the nature of subjects. Possibly the term "brightness" could fairly describe that inherent difference.

If the data in Figure 2 are turned into percentages and arranged separately for bright and dull in the

TABLE 9

AVERAGE PERCENTAGE OF GAIN BY DULL AND BRIGHT PER IDEA AS ARRANGED IN FIGURE 2

Dull Low	Dull Middle	Dull High	Bright Low	Bright Middle	Bright High
1.1	1.75		2.4	3.25	
1.5	2.5	1.83	1.5	2.17	1.83
	3.33	3.33		1.33	1.00

same way that the data are presented in Figure 2, numerical differences can be seen, as in Table 9.

Summarizing the relationships indicated by study of improvement, it appears that, in regard to 6 ideas, the dull gained more than the bright, explained by the fact that the bright had high first scores making much gain impossible; that in regard to 13 ideas, the bright gained more than the dull, explained by the fact that the nature of these ideas was more difficult than that of the other ideas of the selection; and that in regard to 29 ideas, the arithmetical gains were about the same, explained by the fact that these ideas had no special inherent difficulty.

CONCLUSION 7

The bright subjects gained more than the dull on the more difficult ideas.

CONCLUSION 8

The dull and the bright gained about the same on the easy ideas.

A proper question to raise at this point is whether or not the same conclusion will be reached from a comparison of bright-nine and dull-twelve groups,

children of about the same mental age differing about 30 points in IQ. In such a study of the data of this experiment, the reader is reminded that these two groups are but one-half the size of the combined groups of bright and dull upon which the above conclusions regarding improvement were made. This smaller size gives opportunity, of course, for chance factors to affect score differences more strongly.

From the data of Tables 10 and 11, it is found that in the cases of 23 ideas, the bright-nine subjects gained more than the dull-twelve, while in 15 cases the reverse was true. Table 10 lists these ideas, and gives the rank position of each according to Table 4*a*. The important facts here shown are that 14 of the 23 ideas, or 61%, on which bright-nine gained more than dull-twelve, are in the lower half of the rank-order list of ideas; that the average difference in gain per idea is 1.4; that this difference is made up of an average of .5 of a point lower score by bright-nine than by dull-twelve on the first scores, and an average of .9 of a point higher score by bright-nine than by dull-twelve on the last scores; that, as to last scores, in 13 cases of the 23 ideas the bright-nine are higher than those of the dull-twelve; in 8 cases the same; and in two, lower; and that the dull-twelve make a total score at the last which is 84% of the bright-nine score.

The important facts regarding the 15 ideas shown in the tables for which the dull-twelve gained more than the bright-nine are that 3 of the ideas, or 20%, are in the lower half of the rank-order list; that the average difference per idea is 1.5; that this difference is made

TABLE 10

COMPARISON OF GAINS MADE BY BRIGHT-NINE AND DULL-TWELVE

(23 cases bright-nine better than dull-twelve)
(15 cases dull-twelve better than bright-nine)

Idea	Rank Position	Idea	Rank Position
1	13	3	17
2	18	7	19
4	14	9	30
14	1	10	5
15	9	11	6
16	4	12	8
17	37	13	3
18	42	19	23
24	25	20	7
25	34	21	24
27	31	22	33
30	36	26	15
31	43	28	48
33	39	32	40
34	44	45	28
35	43		
36	46		
37	47		
39	11		
40	10		
41	16		
43	32		
47	29		
Percentage in lower half of rank-order list	61		20

up of an average of .8 or a point lower score by the dull-twelve than by the bright-nine on the first scores and an average of .7 of a point higher score by dull-twelve than bright-nine on the last scores; that in 7 cases of the 15 ideas the last score of the dull-twelve is higher than that of the bright-nine and in 8 cases the same; and that the bright-nine make a total score at the last which is 89% of the dull-twelve score.

In summary terms, it appears that, on the whole the bright-nine did better than the dull-twelve, but

not so much better that the picture is clear cut. The superiority of the bright-nine seems to be in the proportion of cases in which they gain more than the dull-twelve and in their ability to make greater gains on the more difficult ideas. Very evidently other unknown factors than those indicated were involved in the process. These factors cancel the effects of one another less extensively in the smaller groups than they did in the larger groups of combined dull and combined bright subjects.

TABLE 11

COMPARISON OF AVERAGES AND PERCENTAGES OF DATA IN TABLE 10

(23 cases bright-nine better than dull-twelve)
(15 cases dull-twelve better than bright-nine)

	Bright	Dull	Difference	Bright	Dull	Difference
Av. last scores	5.8	4.9	.9	6.8	6.1	.7
Av. first scores	3.0	3.5	.5	3.9	4.7	.8
Av. gain	2.8	1.5	1.4	2.9	1.4	1.5
Percentage dull-twelve last scores were of bright-nine last scores						84
Percentage bright-nine last scores were of dull-twelve last scores						89

IV

SUMMARY

Conclusions as to the similarity found among groups in regard to the nature of difficulty.

1. In general the same ideas were relatively the most difficult for all groups.
2. In general the same ideas were relatively the easiest for all groups.
3. Ideas of location were easier than descriptive or abstract ideas.
4. Position in first or last halves of the selection, or at the beginning or end, did not seem to be of importance.
5. Part of speech may be a factor as far as conjunctions are concerned, but does not appear to be so in the case of nouns, verbs, and adjectives.

Conclusions regarding factors having differing effects among the subject groups.

1. Abstractness is relatively easier for bright than for dull children.
2. Sense of coherence and boredom are more effective with bright children.
3. Improvement is arithmetically greater for bright than for dull children on the more difficult ideas, but about the same on the easier ideas.

These differences are not certain. The data suggest them.

Difficulty, as a term describing differences in scores of achievement or of improvement over a period of

practice, embraces a compound of many factors. One of these factors is probably the difficulty in the nature of the small concept itself. Others may be interest, attention, and other subtle relationships among the items making up a series of ideas, and probably equally subtle relationships to the mental organization of each individual. With the ideas used in this experiment, the effect, in terms of averages, of those factors, other than the nature of the difficulty of each idea, is apparently very much the same for both bright and dull, younger and older.

REFERENCE

1. WILSON, F. T. Learning of bright and dull children. *Teach. Coll. Contrib. Educ.*, 1928, No. 292. Pp. 56.

UNE COMPARAISON DE LA DIFFICULTÉ ET DE L'AMÉLIORATION DANS L'APPRENTISSAGE À REPRODUIRE UN MORCEAU DE DESCRIPTION CHEZ LES ENFANTS TRÈS INTELLIGENTS ET PEU INTELLIGENTS

(Résumé)

On a donné de la pratique individuelle à quatre groupes choisis composé chacun de quinze enfants dans l'apprentissage de la reproduction d'un morceau descriptif à cinq jours successifs. Deux des groupes ont été composés d'enfants âgés de neuf ans, l'un des groupes ayant des Q.I. de quatre-vingts à quatre-vingt-dix, l'autre ayant des Q.I. de cent dix à cent vingt. Les deux autres groupes ont été composés d'enfants âgés de douze ans et chacun de ces groupes a eu, respectivement, les Q.I. élevés et peu élevés des enfants âgés de neuf ans. Le morceau pratiqué a été la description assez peu intéressante d'une jeune fille. On l'a lu aux sujets chaque jour et immédiatement après il a été reproduit oralement par eux. On a noté les idées reproduites par les sujets. L'étude montre le rendement moyen de chaque groupe pour chacune des diverses "idées" du morceau, surtout à l'égard de la difficulté relative des idées.

Les résultats pour la reproduction correcte ont indiqué qu'avec une exception toutes les idées du morceau ont été de la même difficulté relative pour chacun des quatre groupes. L'exception a été celle d'une idée décrite comme on ne peut plus abstraite. En comparaison des autres idées les enfants très intelligents âgés de douze ans ont beaucoup mieux réussi avec cette idée-ci que n'importe quel autre groupe. Il paraît aussi de ces résultats que les idées de position aient été plus facilement reproduites par tous les groupes que les idées de description.

On a fait aussi une comparaison des groupes à l'égard de l'amélioration pendant les séries de pratique. À cet égard il paraît que les enfants très intelligents aient montré plus de progrès relativement que les moins intelligents avec les idées donnant les résultats les moins élevés, c'est-à-dire, les idées les plus difficiles. D'autre part il paraît que les enfants les moins intelligents aient montré le même progrès relatifs que les très intelligents avec les idées plus faciles.

L'étude des résultats de l'enquête suggère que les facteurs importants d'une tâche comme celle de cette expérience se trouvent dans la signifiance de l'expérience plutôt que dans le caractère unique des individus qui en ont fait partie. Il est sorti une idée centrale dans les reproductions de tous les groupes, et autour de cette idée d'autres idées en rapport se sont formées. La synthèse suit la même forme générale pour tous les groupes, jeunes et moins jeunes, intelligents et moins intelligents. La similarité des expériences basiques de la vie, sur lesquelles se repose la compréhension du langage, pourrait expliquer ce fait en grande partie.

Outre les suppositions citées ci-dessus, les données suggèrent qu'il s'est trouvé d'autres facteurs ayant rapport au tempérament et influant sur les réactions des enfants. Surtout, les facteurs de l'intérêt et de l'ennui semblent présents. Il est possible que ces facteurs aient été caractéristiques de certains des groupes. Ils paraissent plus effectifs chez les sujets les plus âgés et les plus intelligents. Quand cela serait vrai, cependant, ces facteurs aussi seraient de beaucoup le résultat de l'expérience. C'est de l'expérience que les valeurs relatives se développent et les intérêts deviennent distinctifs. L'expérience des plus âgés et des plus intelligents a plus développé leurs intérêts.

WILSON